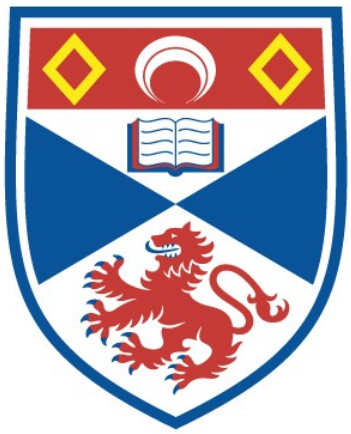
IN SILICO MODELLING OF IN-HOST TUBERCULOSIS DYNAMICS: TOWARDS BUILDING THE PERFECT PATIENT

Michael John Pitcher

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

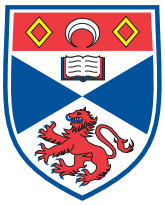


2020

In silico modelling of in-host tuberculosis dynamics: towards building the virtual patient

by

Michael John Pitcher



University
of
St Andrews

This thesis is submitted to the

University of St Andrews

in partial fulfilment for the degree of

Doctor of Philosophy

September 2019

Abstract

Tuberculosis (TB) accounts for over 1 million deaths each year, despite effective treatment regimens being available. Improving the treatment of TB will require new regimens, each of which will need to be put through expensive and lengthy clinical trials, with no guarantee of success. The ability to predict which of many novel regimens to progress through the clinical trial stages would be an important tool to TB research. as current models are constrained in their ability to reflect the whole spectrum of pathophysiology, particularly as there remains uncertainty around the events that occur.

This thesis explores the use of computational techniques to model a pulmonary human TB infection. We introduce the first *in silico* model of TB occurring over the whole lung that incorporates both the environmental heterogeneity that is exhibited within different spatial regions of the organ, and the different bacterial dissemination routes, in order to understand how bacteria move during infection and why post-primary disease is typically localised towards the apices of the lung.

Our results show that including environmental heterogeneity within the lung can have profound effects on the results of an infection, by creating a region towards the apex which is preferential for bacterial proliferation. We also present a further iteration of the model, whereby the environment is made more granular to better understand the regions which are afflicted during infection, and show how sensitivity analysis can determine the factors that contribute most to disease outcomes.

We show that in order to simulate TB disease within a human lung with sufficient accuracy, better understanding of the dynamics is required. The model presented in this thesis is intended to provide insight into these complicated dynamics, and thus make progress towards an end goal of a virtual clinical trial, consisting of a heterogeneous population of synthetic virtual patients.

Candidate's Declaration

I, Michael John Pitcher, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 32000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2015.

I received funding from an organisation or instituiton and have acknowledged the funder(s) in the full text of my thesis.

Signature of Candidate: .....

Date: September 2019

Supervisor's Declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Signature of Supervisor: .....

Date: September 2019

Permission for Publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Michael John Pitcher, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy
No embargo on print copy.

Electronic copy
No embargo on print copy.

Signature of Candidate: .....

Date: September 2019

Signature of Supervisor: .....

Date: September 2019

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Michael John Pitcher, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Signature of Candidate: .....

Date: September 2019

Nid wy'n gofyn bywyd moethus,
Aur y byd na'i berlau mân:
Gofyn wyf am galon hapus,
Calon onest, calon lân.

Calon lân yn llawn daioni,
Tecach yw na'r lili dlos:
Dim ond calon lân all ganu
Canu'r dydd a chanu'r nos.

'Calon lân' by Daniel James / John Hughes

Dedicated to the memory of Elizabeth Ann Lloyd

Acknowledgements

This work was supported by the PreDiCT TB consortium (IMI Joint undertaking grant agreement number 115337, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EF-PIA companies' in kind contribution.)

This work would not have been possible without the help of a multitude of people. Firstly, I would like to thank Prof Simon Dobson, my primary supervisor, for his support, guidance and willingness to listen to me awkwardly try to explain the medical side of this project. His insight and advice have proven invaluable and I couldn't have wished for a better mentor through this project. I'm also fortunate enough to have had input from multiple supervisors during this PhD, and I would like to thank Prof. Stephen Gillespie for his insightful medical knowledge and enthusiasm for my work and to Dr Ruth Bowness for her mathematical insights and for her motivation that kept me going when things got difficult.

I would also like to thank various other contributors who have assisted in publications and other work connected with this project. These include Dr Derek Sloan, Prof Mark Chaplain, and Dr Raluca Eftimie. I would also like to thank the various research groups I have been a part of, whom have provided me great feedback on the work I have presented to them. These include the Infection and Global Health Division of the School of Medicine, and the School of Computer Science's Systems Research Group and Health Informatics Group.

On a personal note, I wouldn't have been able to make such a big life

change and come to St Andrews without the support of those close to me. I'd like to thank my parents, Lynn and Victor, and my brother Matthew, for their constant love, support and encouragement leading up to this and throughout. I'd also like to thank my friends Dave Humphreys and Jo Baker for their encouragement to pursue this opportunity and for their friendship, which has always helped me when I've needed it most. I would also like to thank the friends I've made along the way, including (but not limited to) Praj, Chris, Luke, Frances, Martin, Andrew, Marian, John, Naomi, Scott, Lei and many more. I'd also like to thank Klaus and Ryoko for being so welcoming and supportive.

And finally, I can't express how grateful I am to my partner, Akika Kalde. Without her constant love and encouragement, none of this could have been possible. You make me a better person and for that, I'll always be grateful. I love you.

Michael John Pitcher
St Andrews
September 2019

Contents

Glossary

aerobic Requiring oxygen. 18

alveolar dead space The volume of air inhaled into the lung which does not subsequently take part in oxygen exchange, due to either not reaching the alveoli or reaching alveoli that are inadequately perfused. 31

alveolus small air sac at the end of the bronchial tree which allows for gas exchange. 18, 26, 29, 66, 91

apoptosis The normal and controlled death of a cell. 22, 23, 76

bronchopulmonary segment Residing within a cell. 65, 69, 72, 90, 142

caseum The central necrotic tissue within a granuloma, named for its cheese-like appearance. 24, 26, 34, 95

chemokine Cytokines that induce chemotaxis in cells. 39, 40, 96

chemotaxis Movement of an immune cell driven towards ro away from a chemical gradient. 96

cytokine Chemicals secreted by immune cells which cause other cells to behave differently. 22, 40, 41

extracellular Residing outside of a cell. 39, 74

granuloma A structure of immune cells, fibrosis and tissue inflammation that contains infection and stops the spread of bacteria. 20, 23–26, 137, 140

haematogenous Relating to the blood. 28

hypoxia Being in an inadequate oxygen supply in order to function. 17

intracellular Residing within a cell. 39

lower respiratory tract The structure of the human airway starting from the trachea and extending through the bronchial tree to the terminal alveoli. 18

lymphocyte An immune cell originating within the lymphatic system. 31

morphology The physical form and shape of an object, e.g. a organ in the body. 148

necrosis The premature death of a cell due to injury. 22, 24, 26

pathogenesis The manner in which a disease develops. 25, 26

pathology The study of the causes and effects of disease. 13, 21, 28, 30

pathophysiology Combination of pathology and physiology - the impact that disease has upon the physiology of the body. 4, 25, 26, 35, 67, 95, 139, 141

phagocyte A type of immune cell whose purpose is to patrol tissue and ingest and destroy any foreign bacteria or particles. 22, 74

phagocytosis The process of ingestion of foreign materials, such as bacteria, by phagocyte cells. 74

phenotype The observable physical properities of an organism. 5, 32, 41, 74

physiology Branch of biology dealing with the function of the living organisms. 28, 31, 148

pulmonary Affecting the lungs. 8

zoonotic (Of disease) contracted from animals. 18

Acronyms

M. africanum *Mycobacterium africanum*. 16

M. bovis *Mycobacterium bovis*. 16, 18

M. tuberculosis *Mycobacterium tuberculosis*. 2, 4, 16, 17, 20, 22, 23, 27, 28, 30, 31, 33, 34, 39, 41, 68, 74

ABM agent-based model. 39–41

APC antigen-presenting cell. 23, 27, 41, 75, 132

BCG Bacillus Calmette–Guérin. 14

DC dendritic cell. 23, 41

eFAST extended Fourier amplitude sensitivity test. 117, 119–124, 135, 138, 146

HIV human immunodeficiency virus. 15, 16, 20, 25, 41, 108

LHS Latin Hypercube Sampling. 114–116, 121, 130

MDR-TB Multidrug-resistant tuberculosis. 33

NHP non-human primate. 3

ODE ordinary differential equation. 40, 41

PD pharmacodynamic. 40, 149

PDE partial differential equation. 39–41

pdf probability density function. 114

PK pharmacokinetic. 40, 149

PRCC partial rank correlation coefficient. 116, 117, 120–123, 125, 129–135, 137, 144

SA Sensitivity Analysis. 113, 115, 119–121, 124, 125, 128, 135, 136, 138

SEIR Susceptible-Exposed-Infectious-Recovered. 43, 56

SIR Susceptible-Infectious-Recovered. 43, 59

SIS Susceptible-Infectious-Susceptible. 43

TB tuberculosis. xxvii, 1–6, 8–10, 13–16, 18–21, 24–28, 30–35, 37–42, 49, 56, 61, 63, 64, 66, 67, 77, 85, 86, 89, 91, 95, 107–109, 111, 112, 120, 136, 137, 139–144, 146–149

TST tuberculin skin test. 20

UA Uncertainty Analysis. 112–114, 119–121, 125, 130, 135

WHO World Health Organization. 1, 2, 32

XDR-TB Extensively drug-resistant tuberculosis. 33

List of Figures

List of Tables

Chapter One

Introduction

In which the motivation behind the thesis is explored, mechanistic modelling is discussed and research questions and thesis structure are outlined

1.1 Motivation

This thesis explores the ways in which computational models of disease can contribute to the understanding of observed pathology, and create insights to guide future medical exploration and treatment development. We look at one disease in particular: tuberculosis (TB).

1.1.1 The eradication of TB

TB is a bacterial-based infectious disease. Infection may be naturally resolved by the host's immune system but can also be fatal, and TB is classified by the World Health Organization (WHO) as the leading cause of death from a single infectious agent, with an estimated 10 million people contracting TB in 2017 and over 1.4 million deaths being attributed to the disease [245]. It is estimated that approximately one quarter of the world's population is infected with some form of the disease, often in a latent, asymptomatic form [115]. TB has been present in human societies for thousands of years [53] and has claimed an incalculable number of lives. Since the introduction of the first vaccine for TB in the 1920s and the first anti-tuberculous drug, streptomycin, in the 1940s, TB incidence has declined rapidly in developed regions such as western Europe, Japan

and North America [76], to the point where it is often considered, in those regions, to be eradicated and an 'ancient' disease. But this does not reflect the current global situation, with high prevalence levels occurring in India, China and sub-Saharan Africa.

The WHO recommends a standard treatment regimen that can be used to treat drug-susceptible forms of TB, which make up the majority of cases worldwide [246]. When correctly adhered to, this regimen should eliminate the bacteria within the patient and thus cure them, which raises what seems like a fairly obvious question:

> If we have an effective treatment regimen that works to cure TB, why has the disease not been eradicated globally?

The answer is not straightforward. One aspect that contributes to the continuing global burden of TB prevalence and mortality is adherence. The standard regimen requires the patient to commit to 6 months of daily chemotherapy using multiple drugs, all of which have adverse side-effects. This duration is very long but also necessary - shorter regimens have been shown to have an unacceptable relapse rate [79]. It is believed that this long time-frame is necessary in order to kill the entire population of *Mycobacterium tuberculosis* (*M. tuberculosis*) bacteria which cause infection, as these mycobacteria are better able to withstand chemotherapy due to them becoming metabolically dormant [142]. The combination of a long duration and toxic side-effects can contribute to patient non-adherence, with patients being overwhelmed by the multiple drugs or discouraged by the negative side-effects from completing treatment [58]. Furthermore, patients may start to feel better before treatment has concluded and thus not complete the full regimen as they believe they are completely cured.

Non-adherence has serious implications. At the individual level, failure to complete treatment may result in some bacteria remaining alive in the patient, even in the absence of any outward symptoms of infection; these bacteria then act as a reservoir for future relapse and further disease. Further compounding this is the phenomenon of drug resistance: exposure to insufficient quantities of a drug can lead to the remaining

organisms (i.e. those not destroyed by it) developing resistance to that drug [89], and some bacteria may become multiple-drug resistant [61, 253]. This emergence of resistance is a danger to the patient as it increases the chance of relapse and of treatment failure. But, far worse than this, drug-resistant strains of disease pose a major threat to the global control of TB [89, 175], as they are harder to treat and require the use of more expensive and more toxic second-line drugs. Improving patient adherence is, therefore, of paramount importance but efforts to improve adherence through direct observation of patients taking medicine have been shown to have no impact on clinical results [129]. It has been proposed that rather than trying to force adherence to the standard regimen, it may be more sensible to make the standard regimen easier to adhere to. This could possibly be done by reducing the length [64, 253], which would require new regimens involving different drugs than the current gold standards.

1.1.2 Creating shorter treatment regimens

Creating new treatment regimens for any disease is a difficult and expensive process. For TB, many new possible regimens exist [253], but these must all be evaluated in extensive (and costly) clinical trials in order to prove their effectiveness, with no guarantee of a successful result: three recent phase III clinical trials, attempting to shorten treatment time to four months, proved unsuccessful [90, 127, 162].

The ability to predict in advance which of many multiple possible new regimens is most likely to be successful when put to clinical trial would be of great benefit to the TB research community. Current models used to do this are typically either *in vitro* laboratory experiments and *in vivo* animal experiments. These models are key to identifying potentially useful agents prior to human administration, but it is important to acknowledge that the benefit of these analyses is dependent on how closely the experimental model approximates the pathophysiological conditions of human disease: *in vitro* laboratory models only offer a narrow view of a single aspect of the complex inter-play of dynamics of bacteria and

immune cells that occur during infection, and whilst a number of *in vivo* animal models exist, including mice, zebrafish and non-human primates (NHPs) [99], each of these models either does not accurately reflect the pathophysiology found in humans and/or has ethical concerns [98].

Statistical models of TB are also involved in the clinical trials process, often predicting individual treatment response based on the rate of bacterial clearance in the patient, based on data collected from sputum samples. Whilst these models are a crucial part of this process, they are constrained by the data upon which they are built, especially if the data available are sparse or do not appropriately reflect the full conditions of the disease. The analytical outputs of TB statistical models have never been robustly validated as predictors of long-term clinical outcome. For example, the relationship between sputum culture conversion and final patient outcome is imperfect and becomes unreliable when looking at data from individual patients. The REMoxTB phase III trial unsuccessfully trialled a new 4-month regimen containing moxifloxacin [90], and a recent analysis of this demonstrated that there was a small sub-set of patients who relapsed after treatment even though their sputum samples rapidly became culture-negative, indicating that the clinical outcome could not be based solely on this data [189].

It is possible that the *M. tuberculosis* cells which are critical for defining patient relapse may not actually be present in the sputum [188]. They may in fact be sequestered in isolated sites within the lung [122] such as the closed lesions described by Canetti [30], and thus be undetectable in sputum samples. Understanding where bacteria reside during infection is important: recent work has suggested that heterogeneous penetration by components of drug regimens into pathological TB lesions can result in insufficient levels of some drugs to reach the minimum inhibitory concentration (MIC) for bacterial elimination [196]. This may have contributed to relapse in the experimental arms of the REMoxTB trial.

M. tuberculosis can exist in a state where many of its metabolic processes are down-regulated and this is often associated with the presence

of lipid body in the mycobacterial cell [83, 104, 171]. These cells are postulated as a cause of relapse in treated patients [142, 189] as the presence of lipid inclusions is associated with a significant increase in the amount of antibiotic needed to destroy the bacteria, meaning the they are more likely to survive the treatment process and go on to cause future relapse [92, 104]. Many possible influencing factors have been postulated as the driving force behind this switch to 'dormancy', but the exact determinants are uncertain [104, 142]. As current TB statistical models are typically based only on standard drug susceptibility tests that measure lipid-poor cells that are phenotypically more susceptible to antibiotics, they do not account for the whole spectrum of bacterial phenotypes.

Therefore, in order to build more effective statistical models, we require mechanisms that capture the biological processes and environmental conditions within the host which have the greatest influence on disease progression. Mechanistic models can provide this detail.

1.1.3 Mechanistic modelling of disease

Mechanistic modelling involves the simulation of events as they occur in the real world, through mathematical or computational means. The overall system is broken down into constituent parts, each of which is simulated: the overall system behaviour emerges as a result of the combination of interactions of the individual components. In this manner, complex systems such as those seen in nature can be achieved through describing just the simplistic individual events that comprise them [243]. Although these models represent an abstracted view of the complexity involved in the real-world systems they simulate, they can still provide insight into those events and parameters which influence important outcomes. This abstraction affords us the power to describe events even in cases where laboratory diagnostic tools or treatment monitoring tools are unable to measure them.

As seen in Figure 1.1, mechanistic model development is typically an

iterative process [25]. An initial model is created based on a theoretical description of the underlying real-world dynamics. This model then enters a period of iterative refinement along with testing against available data until the model is deemed to adequately describe the data. Once this has been achieved, the predictions created by the model are then validated against independent experiments or trials. If the predictions are validated, the model's parameters and dynamics may provide important insights that further our understanding of the real-world phenomena of the system being studied, advancing our understanding. These advances can be fed into further models or can direct future experiments or trials (or both).

Mechanistic models are inexpensive to produce compared to a clinical trial or in vitro experiment, fast (obtaining results in hours rather than weeks or years), and easily adaptable to account for new data [98]. Furthermore, the development of the mechanistic model can be as important to furthering our understanding as the predictive outcomes which the model provides. Each parameter of the model must have a biological basis and the process of parameterising these models can identify important knowledge gaps, thus serving as a guide to direct new research.

In silico models have long been applied to the study of TB, but these have predominantly focused on the spread of TB epidemics [229]. Whilst these models are essential in determining effective intervention strategies [111, 140, 170], epidemic modelling alone will not be enough if we wish to reduce TB mortality worldwide as it only allows the reduction in transmission, not the removal of the source of infection. TB *in silico* modelling needs greater focus on the in-host dynamics in order to make the gains seen within other fields such as cancer research [103]. This is a relatively new concept for TB and has only recently started to gain momentum (see [132] for a comprehensive review of in-host TB modelling efforts).

Simulation of within-host TB dynamics could occur at many scales. Lesional level models simulate the interactions between host and bacteria on a small-scale, looking at the individual factors that impact the development and progression of a single lesion of infection. At a larger scale,

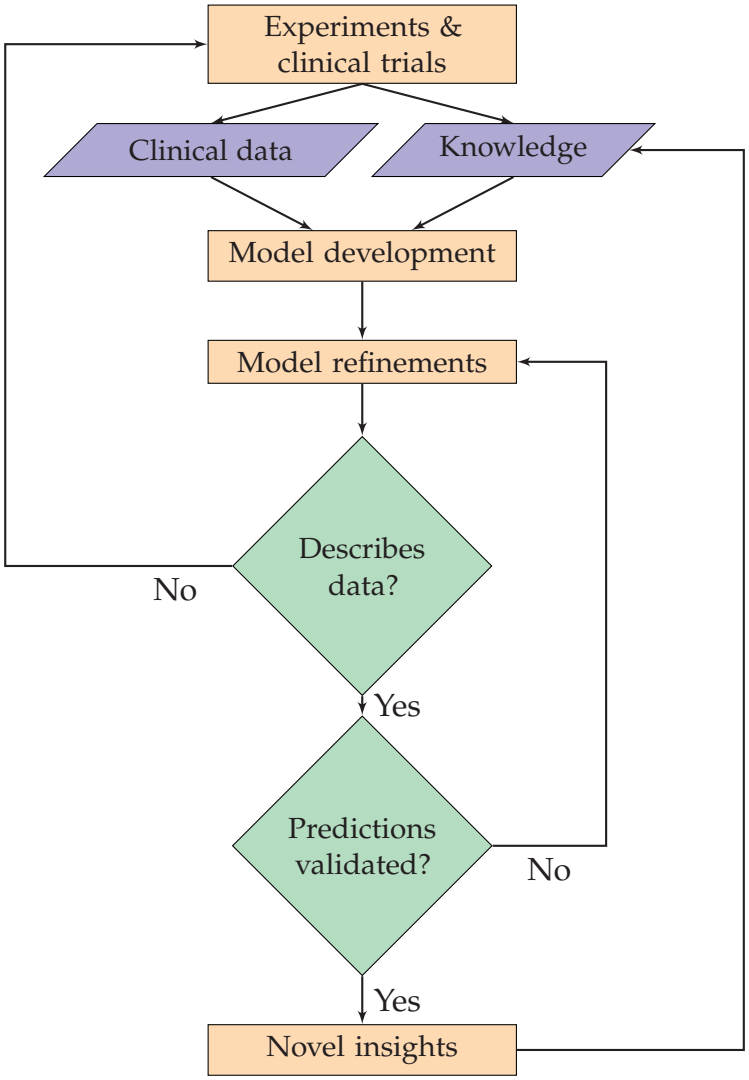


Figure 1.1: The iterative process of mechanistic model development. Laboratory experiments provide patient demographic and clinical data, as well as furthering our understanding of the real-world system. Our current understanding provides the system dynamics of the model and quantifiable real-world data provide the parameters. The model may be refined, and this process of development may highlight any missing knowledge, which then guides further experiments or trials to collect the relevant data. Once simulation predictions are obtained, they must be validated against independent results. If the predictions cannot be validated, the model must be refined. If predictions are validated, the dynamics of the model provide insight into the processes driving the observations and thus furthers our knowledge (adapted with permission from [25]).

whole-organ or whole-body models can explore how variance of biology within and between hosts can impact both the disease and its treatment.

Currently, most mechanistic modelling of in-host TB dynamics has focused on a lesional scale and has provided valuable insights into the factors that contribute to individual lesions. However, during an infection, many lesions can develop in a variety of spatial locations within the lung. Therefore, larger scale models looking at the whole lung and whole body are needed, investigating how individual lesions form in specific locations within the lung and later coalesce to form cavities for bacterial dissemination between people.

1.2 Research questions

In this work, we seek to answer the following research questions:

- To what extent can we use computational models of whole organ scale to provide insight into the dynamics that occur during TB infection within the lungs?
- What environmental factors within the lung contribute to the specific localisations of TB during different stages of the disease?

1.3 Contributions

This thesis makes the following contributions to the field in order to answer the research questions in section 1.2:

- A framework for the creation of networked metapopulation models which incorporate environmental heterogeneity
- A whole organ model of a pulmonary TB infection that incorporates the environmental heterogeneity found within the lung to determine the role it plays upon disease progression
- A software framework to efficiently perform sensitivity and uncertainty analysis

- Insight into biological factors that contribute to TB progression

1.4 Publications

The following publications were published based on the work conducted in this thesis:

- M. Pitcher et al. 'A Network-Based Metapopulation Model to Simulate a Pulmonary Tuberculosis Infection'. In: *The 6th International Conference on Complex Networks and Their Applications BOOK OF ABSTRACTS*. Lyon, 2017, pp. 61–63

- M. J. Pitcher et al. 'A spatially heterogeneous network-based metapopulation software model applied to the simulation of a pulmonary tuberculosis infection'. In: *Applied Network Science* 3.1 (Dec. 2018), p. 33. ISSN: 2364-8228. DOI: 10.1007/s41109-018-0091-2 (comprised of the work detailed in Chapters 3 and 4)

The following publication is in pre-print:

- M. J. Pitcher et al. 'Modelling the effects of environmental heterogeneity within the lung on the tuberculosis life-cycle'. In: *bioRxiv* (2019). DOI: 10.1101/2019.12.12.871269

1.5 Thesis structure

This thesis is structured as follows:

- Chapter 2 provides an overview of the biology, pathology, immunology and therapeutics of TB which are relevant to this thesis, paying particular attention to the distinct localisations that occur during different stages of TB infection.

- Chapter 3 describes the development of a framework for the creation of networked metapopulation models. This framework allows for a network of connected patched with heterogeneous environmental attributes to be created, along with a systems of dynamics that

occur at different rates depending on these attributes. This chapter also details the verification tests performed to ensure accuracy of the system.

- Chapter 4 describes the creation of a whole-lung model of TB, *TBMetapopPy*, using the framework of the previous chapter, that incorporates the environmental heterogeneity within the lung that is believed to contribute to the distinct apical localisation during post-primary disease. Here, we present the results of our initial simulations to show how differentials in environment within the lung can lead to the creation of regions of preferential bacterial growth.

- Chapter 5 details the creation of a secondary iteration of the *TB-MetapopPy* model that incorporates a more granular network. This chapter describes the development of a space-filling tree algorithm used to create a network with a much greater number of patches, which is then combined with the known lung heterogeneities to create a new, granular landscape where differentials are based on vertical position. This model is extended to also include post-primary dynamics, thus making it whole life-cycle model of TB. Here, present our results from this model, showing how environmental conditions affect the disease at all stages of the life-cycle.

- Chapter 6 explores the concepts of uncertainty and sensitivity analysis as used for mathematical and computational models. It describes the creation of a framework designed to allow for easy computation of the parameter samples required for these analyses, as well as the automatic model evaluation using these samples in parallel on a remote compute cluster. It presents the results of sensitivity analyses run against the *TBMetapopPy* model, showing how uncertainty in the model outcomes can be attribute to specific model parameters.

- Chapter 7 summarises the content of the thesis, and describes how the contributions made by this work answer our research questions.

We also outline the limitations, both in terms of mechanistic modelling in general and those specific to our work. It also proposes avenues of possible future research arising from the work presented here.

Chapter Two

Background on tuberculosis

In which we review the history of tuberculosis, the pathology of the disease (paying particular attention to the differences in localisation during disease stages), and its treatment

2.1 History of TB

TB is an ancient disease, also known throughout history as 'pthisis' and 'consumption' (due to the disease seeming to 'consume' the host from the inside). A progenitor to the bacteria that currently cause infection has been shown to have existed over 3 million years ago [100]. There is evidence of TB infection stretching back to ancient Egypt (around 1550–1080 BC) [178], and it is believed to have been spread to the Americas over 10,000 years ago with migration from Asia, many years before colonisation by European explorers [54]. Like many other infectious diseases, TB incidence has surged and then receded in a repeating cycle, albeit with time scales much larger than other diseases [53], on the scale of one hundred to several hundred years [21].

By the 1800s, TB had reached epidemic proportions throughout Europe and America. It was during this century that the pathology was first described by René Laennec in 1819, whilst Robert Koch first identified

and described the bacteria causing TB in 1881. Increased scientific study and advancements in medicine and public health systems lead to major progress in fighting the disease and a general reduction of TB incidence in Europe throughout the 20th century, albeit with a sharp increase during the First World War. In 1921, the Bacillus Calmette–Guérin (BCG) vaccine was first trialled on humans, with the vaccine successfully conferring immunity to TB on child recipients. In 1943, streptomycin, the first anti-TB drug, was discovered and further drugs were developed in the following years.

In the twentieth century, the number of cases and deaths rapidly declined in most developed countries, including western Europe, Japan and North America [76], caused by the introduction of new treatments and preventions and a range of complex social and economic changes [95]. However, incidence and mortality have not seen such declines in less-developed countries.

2.2 Current epidemiology and global burden of TB

Despite effective treatments having existed for decades, TB still remains a global health threat today. In 2017, an estimated 10 million people contracted TB, with over 1 million deaths being attributed to the disease [245]. It is estimated that approximately one quarter of the world's population is infected with the disease in a latent form [115].

The distribution of the disease worldwide is uneven, with far greater prevalence in developing countries than in developed countries. Over half of all new cases of TB in 2017 occurred in just five countries (Figure 2.1), while India alone accounted for more than a quarter of the global incidence. Figures 2.2 and 2.3 show the incidence and mortality, respectively, of TB per 100,000 members of the population for every country in 2017. Sub-Saharan Africa and south-east Asia show significantly higher incidence densities than the rest of the world, while sub-Saharan Africa

has a higher rate of mortality per population. There are many sociopolitical causes of this uneven distribution, including poverty levels and human immunodeficiency virus (HIV) rates [43].

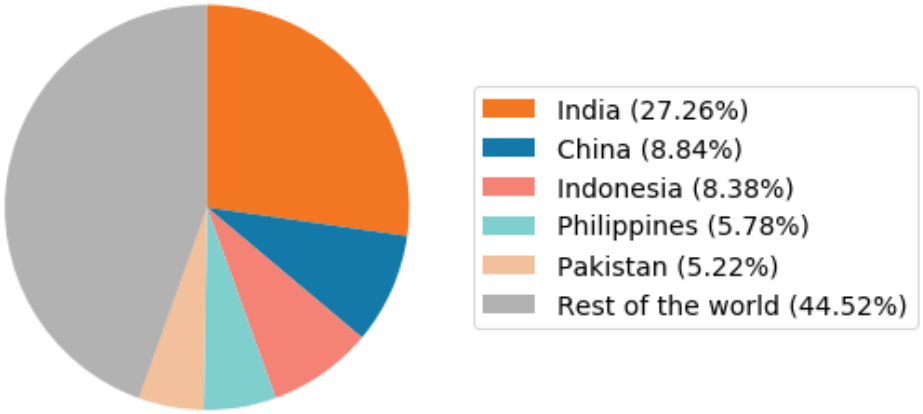


Figure 2.1: Distribution of global incidence of TB in 2017 (top 5 countries with highest values listed). Data taken from [239]

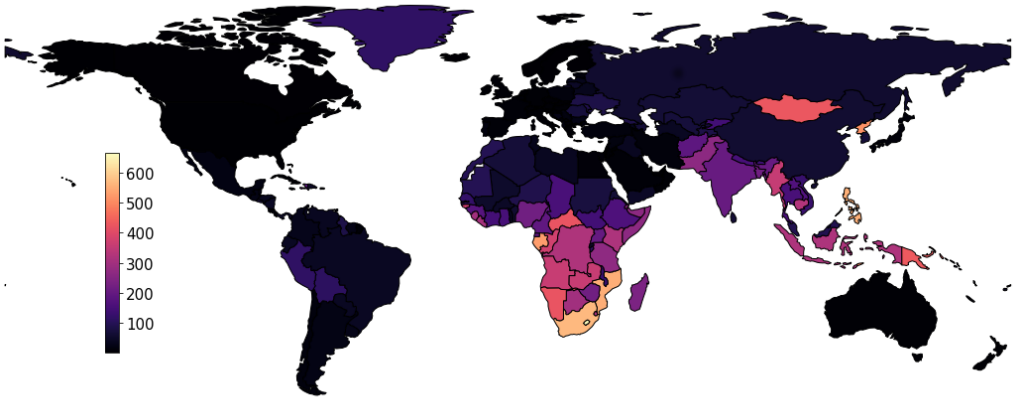


Figure 2.2: Estimated incidence of TB per 100,000 population, by country in 2017. Data taken from [239].

Approximately 9% of the incident TB cases in 2017 came from those co-infected with HIV, with African countries having the highest proportion,

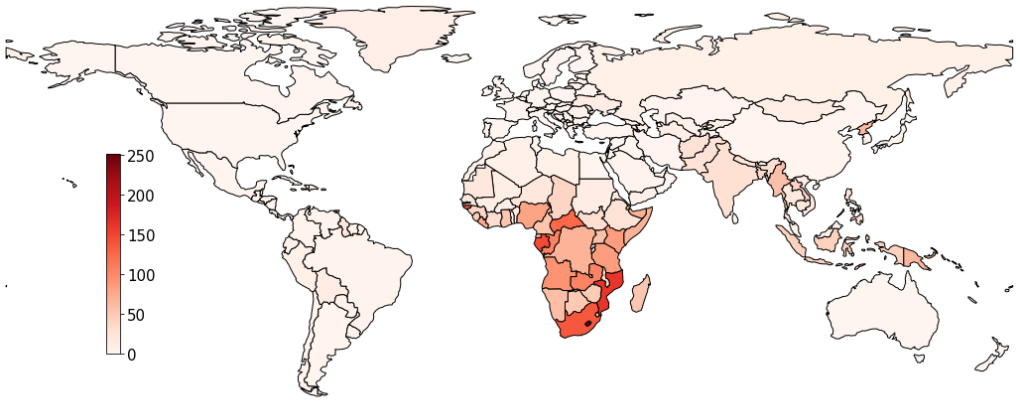


Figure 2.3: Estimated mortality of TB cases (all forms) per 100,000 population by country in 2017. Data taken from [239].

exceeding 50% in parts of southern Africa. The risk of developing TB for those already infected HIV was roughly 20 times greater than the risk for those without HIV [245]. Figure 2.4 shows the positive correlation between HIV rates per population and TB incidence rates per population for 2017, with higher prevalence of HIV resulting in an increase in the number of new cases of TB.

2.3 Cause

TB disease can be found in both humans and animals, and is caused by mycobacteria from the *Mycobacterium tuberculosis* complex. Members of this complex include *M. tuberculosis*, which causes TB exclusively in humans, *Mycobacterium bovis* (*M. bovis*), which can infect humans as well as animals such as cattle and badgers, and *Mycobacterium africanum* (*M. africanum*), a less virulent bacterium similar to *M. tuberculosis* [69]. Mycobacteria are slender, non-motile bacteria and are often highly resistant to antibiotics due to their unusual cell structure and the low permeability of their cell walls [27].

M. tuberculosis, the major cause of TB in humans, is a relatively-slow growing bacterium, with a replication rate of at least 12 hours (for com-

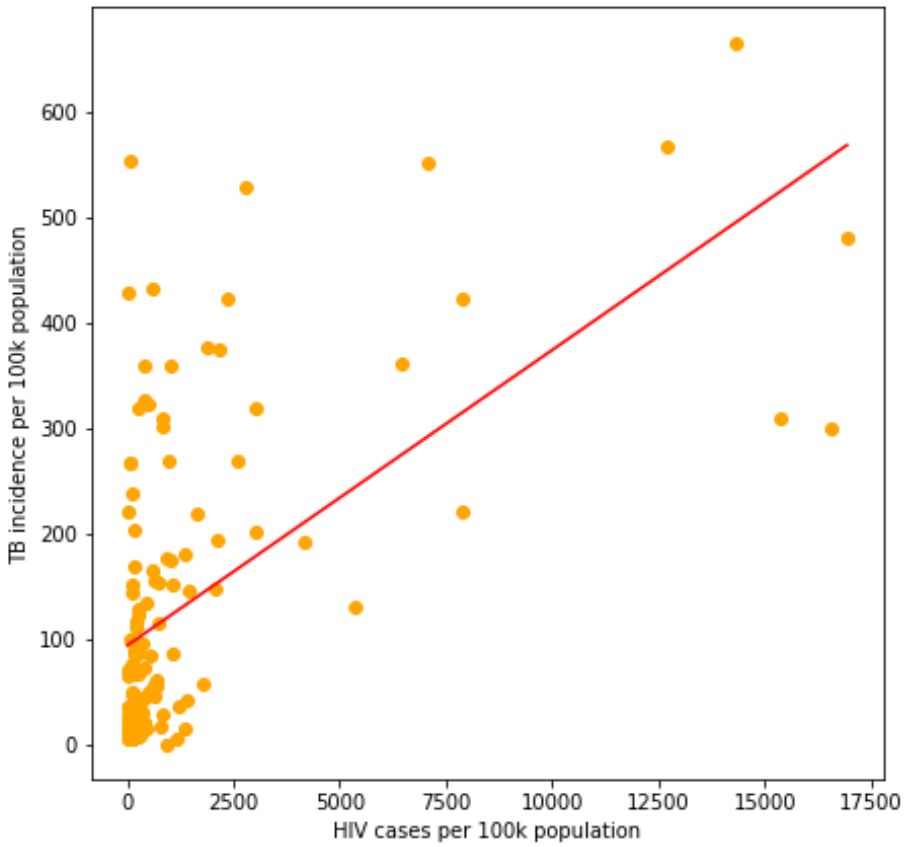


Figure 2.4: HIV cases per 100k population plotted against TB incidence per 100k for year 2017 for every country where HIV data exists. Red line is linear regression performed on data sets (p=1.142e-13). TB data taken from [239], HIV data taken from [238].

parison, *Staphylococcus aureus* has a doubling time of around 42 minutes). *M. tuberculosis* are believed to be able to enter a 'dormant' state, associated with the presence of lipid bodies accumulated by the bacteria [104], whereby they reduce their metabolic activity (lowering their replication rate) and become increasingly resistant to antimicrobial agents [142]. Bacteria in this state can survive within the host for decades [44]. It has been shown that a 'dormant' state can be achieved in a laboratory setting by subjecting the bacteria to stresses such as hypoxia [232] (*M. tuberculosis*

is an aerobic bacteria and thus requires oxygen in order to replicate), nutrient deficiency [144], acidity [216], and antibiotics [117].

2.4 Transmission and initial infection

TB is transmitted between humans primarily through the inhalation of bacteria expelled from the lungs of an infectious individual, although TB can be zoonotic and *M. bovis* strains may be contracted through consumption of unpasteurised cow's milk [50]. Infection can be spread from an infectious person through coughing, talking or other vocal actions, which generate aerosols containing droplets to be expelled from the lower respiratory tract. These droplets are of such a size that they may be immediately respired by other individuals in close proximity, and each droplet may contain bacteria [74], with only a small number of bacteria needing to be inhaled to initiate an infection [33]. Once inhaled, the bacteria pass through the lower respiratory tract eventually landing in the alveoli, where they attempt to establish an infection.

2.5 Disease life-cycle and clinical disease expressions

Once TB infection begins within the lungs, a long series of complex interactions occur between the invading bacteria and the host immune system. Figure 2.5 shows an overview of the TB life-cycle within an individual. Initial exposure is either naturally cleared by the host's defence system or an infection is established within the lung. In approximately 10% of those infected [2], the immune system is incapable of suppressing the bacteria and the infection develops and causes active disease. This is known as 'primary TB' as it originates from the initial infection.

In the majority of cases, the protective immune response is sufficient to keep the infection contained. In a further 10% of cases [2], this eradicates the bacteria from the lungs, but most of those infected will instead become latently infected - whereby the infection remains in the body

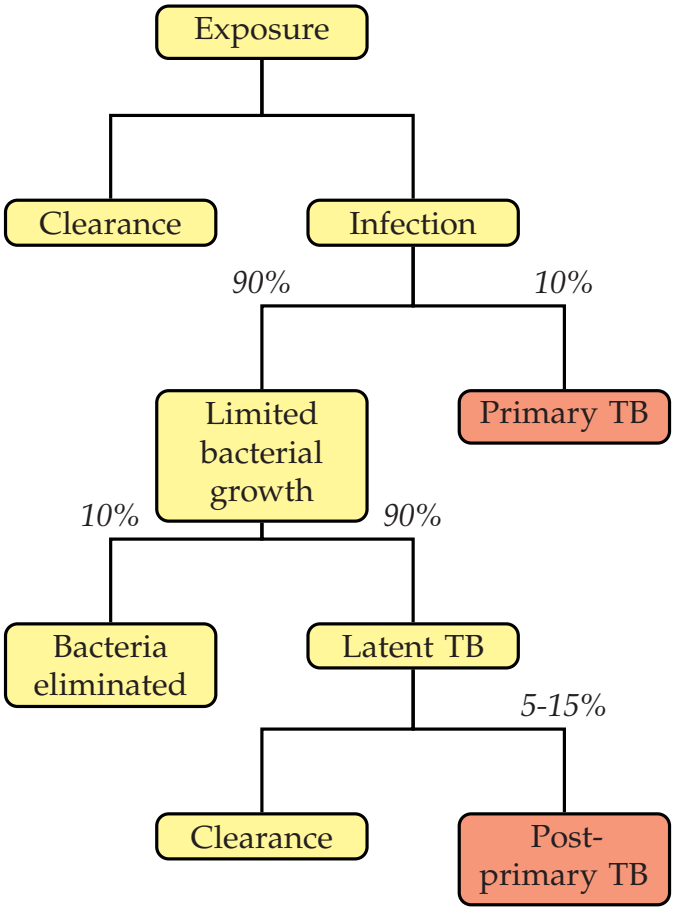


Figure 2.5: Overview of the clinical outcomes of tuberculosis infection. Adapted from [2], values from [2, 245]. Red boxes indicate active, symptomatic infection.

asymptomatically, neither being eradicated nor causing damage and becoming capable of transmission. This period of latency can last for years or decades. Latent infection poses the risk of future active disease, with around 5 to 15% of those latently infected going on to develop active disease in the form of 'post-primary' infection. The exact causes for this re-emergence of disease remain controversial, occurring either due to endogenous re-activation (as the immune system becomes compromised due to old age or co-infection) or due to exogenous re-infection (see section 2.5.4).

Here, we overview the main clinical disease expressions of a TB infection.

2.5.1 Clearance

In rare cases, the innate immune response is strong enough, and the bacterial load is small enough, to completely eradicate infection before an adaptive immune response is established. The evidence that a clearance scenario exists comes from studies that show close exposure to infectious patients does not always lead to a positive tuberculin skin test (TST), indicating that an adaptive immune response has not been triggered, and thus it can be assumed that the immune system has cleared the infection without need of an advanced immune response. [227].

2.5.2 Primary TB

In a small number of cases (approximately 10% [2]) where an infection is established, the immune system is unable to contain the infection and the bacteria present are able to overwhelm the immune cells and cause tissue damage. As progression of TB is heavily dependent on the host immunity, any condition that causes the immune system to become compromised (such as HIV) increases the risk of primary TB [165]. This explains the correlation seen in Figure 2.4, with greater HIV rates increasing the incidence of TB.

2.5.3 Latent TB

In the majority of cases of those infected with *M. tuberculosis* (approximately 90% [2]), the immune system is strong enough to contain, but not eradicate the bacteria, within the granulomas, and thus latent TB is established. Latent TB was defined by Amberson as "the presence of any tuberculous lesion which fails to produce symptoms of its presence" [8]. The definition of latent TB relies solely on the basis of lack of evidence of clinical symptoms, and as such may actually cover a wide variety of patients with differing internal conditions, none of whom display any outward effects of infection [14]. Bacteria within the patient may have entered a dormant phase and thus have reduced replication, or they may be actively replicating but kept at sub-clinical levels. Therefore, latency may be viewed as a 'spectrum' [251].

The classical view of TB pathology is that primary infection typically resolves to latency, with the bacteria numbers contained but not eradicated, and with the initial lesion releasing small numbers of bacteria that are disseminated across the lung via the bronchial tree and the bloodstream, eventually seeding other lesions in the lung [34] (see section 2.8 for more detail on the means of dissemination). This state persists for years or decades, with the bacteria present in the body acting as a reservoir for possible future disease in the form of post-primary TB.

2.5.4 Post-primary TB

Post-primary, or 'secondary', TB accounts for 80% of all clinical TB disease and almost 100% of disease transmission [121]. It can be classified as any form of TB disease that arises after a state of immunity has been established from the initial infection [124]. Thus, post-primary TB must always follow latency.

There are believed to be two routes through which this can occur [13, 69]:

- Reactivation - where infection occurs endogenously, through the reactivation of dormant disease in the lung originating from the initial infection
- Reinfection - where infection occurs exogenously, through the inhalation of new bacteria into the lung

There remains controversy around which of these two causes accounts for the majority of post-primary TB incidence [201].

2.6 Pathogenesis

In this section, we detail the bacteriological, immunological and pathological processes that contribute to the clinical disease expressions outlined above.

2.6.1 Innate and adaptive immune responses

Within the lung, multiple immune cells are present to detect and remove foreign bodies that may interfere with the lung function. The main bulk of these are macrophages, phagocytes which act as sentries: patrolling the environment and ingesting and destroying any invading entities they encounter. When a particle is ingested, it is stored inside the macrophage in a phagosome, a vacuole within the macrophage cell membrane. The macrophage also contains a lysosome, an organelle containing degradative enzymes. During normal function, these two chambers are merged together into a single one called a phagolysosome, exposing any material in the phagosome to the destructive enzymes of the lysosome and thus destroying the material. Whilst *M. tuberculosis* are ingested in this manner, the bacterium has evolved a number of mechanisms to counter its destruction within the macrophage, including prevention of the phagosome–lysosome fusion [161]. By doing so, the bacteria remain present within the macrophage, absorbing nutrients such as iron directly from the immune cell [47], and they begin replicating intracellularly. Excessive replication within the macrophage can cause the immune cell to rupture or 'burst', destroying the cell and releasing its interior bacterial load back into the extracellular space [62], thus allowing for a new cycle of cell infection. Further macrophages are recruited to the site of infection in response to cytokines given off by infected and activated cells there [93].

Rather than allow the bacteria to proliferate within the cell and eventually cause necrosis, the body can trigger apoptosis: the programmed death of the cell [7]. The cell is ordered to destroy itself in an organised manner [136], which is believed to destroy some or all of the intracellular bacteria, as apoptosis of macrophages is associated with a reduction in bacillary viability [183]. Macrophages can also become 'activated' in response to signals from their local environment, with a wide spectrum of possible triggers contributing to this [169]. Activation increases the bactericidal functions of macrophages, allowing them to successfully destroy *M. tuberculosis* bacteria.

Dendritic cells (DCs) also reside in the lung, and these cells also phagocytose bacteria, although they are even less proficient at mycobacterial destruction than macrophages [22]. Instead, the purpose of these cells is to act as a signal to initiate an adaptive immune response [12]. DCs that come into contact with bacteria may maturate, either through contact with antigens (molecules derived from the bacteria) released by necrotic cells or bacteria, or via ingestion of whole bacteria [119]. These cells then migrate from the lung tissue to the adjacent lymph nodes [226] - it has also been shown that macrophages may perform the same migratory function [130].

At the lymph nodes, these antigen-presenting cells (APCs) function by presenting antigens on their surface. Naïve T cells survey these APCs and recognise the presented antigens, which cause the T cells to be primed and convert into an activated state [23]. Once activated, these T cells may migrate back to the lungs, using chemokine signals and homing pathways in order to determine which tissues to migrate towards [9]. There are wide variety of T cells, each of which performs different functions at the site of infection, including activation of resting macrophages and enhancing apoptosis of infected ones. The end result of this is the creation of a cell-mediated immune response, with increased immune activity in the lung intending to suppress infection before it can cause tissue damage. This adaptive immune response takes around 5 to 6 weeks to be initiated, a much longer time-frame than other microbial diseases, possibly caused by *M. tuberculosis* inhibiting the ability of DCs to migrate to the lymph nodes [202].

2.6.2 The granuloma

At the sites of infection, the cell-mediated immune response results in the formation of structures known as granulomas - organised, dynamic arrangements of immune cells designed to contain infection [98] (see Figure 2.6). Infected macrophages compose the centre of the granuloma,

and these are surrounded by layers of infected cells and then non-infected cells and lymphocytes such as T cells. Granulomas consist of a wide variety of cells and are thus extremely heterogeneous [231]. In some cases, the centre of the granuloma may contain a cheese-like substance known as caseum, formed from the necrosis of infected immune cells. The initial lesion formed by TB infection is known as the 'Ghon focus', with a combination of this lesion and an infected lymph node known as the 'Ghon complex'. Whilst it is believed granulomas are a necessity for the host to constrain infection, it is also possible that these structures are actually beneficial for the bacteria, allowing them to persist for long periods of time [204, 217]. The granulomas involved in latent TB often

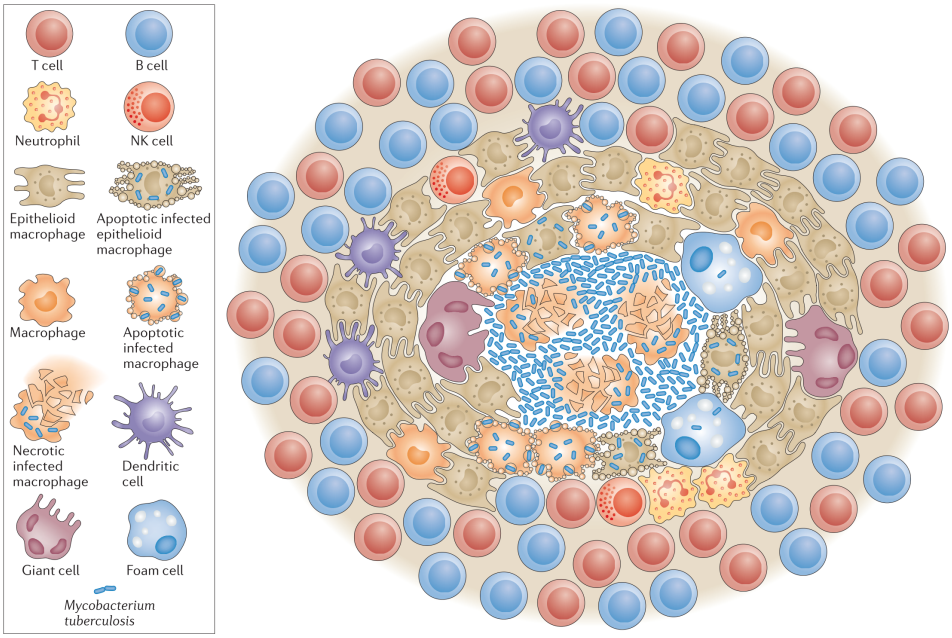


Figure 2.6: Diagram of a granuloma (from [199])

become fibrous in nature, with a fibrotic cuff surrounding the cells and physically isolating infected tissue to prevent disease spread [231]. The exact processes that follow remain controversial. In many cases, the caseum that makes up the centre of the lesion begins to calcify [38]. The majority of these calcified lesions destroy any bacteria contained within them [31].

2.7 Differences between primary and post-primary disease

Whilst a number of contemporary sources view primary and post-primary forms of TB as being similar, grouping them together under a simple label of 'active' disease, it is important for our purposes to make the distinction between the two, as each has differing manifestations of pathophysiology, immunology and localisation.

2.7.1 Differences in localisation

A classical paradigm of TB pathogenesis [29] is that primary infection forms a granuloma that caseates but stabilises, and this granuloma then remains in a quiescent state, holding the bacteria at bay until a weakening of the immune system (e.g. through old-age or contraction of HIV) causes the structure to break down, releasing bacteria and thus allowing them to replicate, causing post-primary TB. This would suggest that both forms of TB occur at the same location. But there is a wealth of evidence to contradict this.

Primary lesions are found in the lower regions of the lung and the surrounding lymph nodes. Post-primary TB is always associated with cavitation [30], and these cavities are found within the apical regions of the lung [20, 67, 122]. This suggests that if primary and post-primary disease originate from a granuloma, then these entry and exit granulomas must be distinct as they occur in different regions of the lung. This would imply a dissemination event - whereby some of the bacteria that make up the initial infection are able to migrate from the lower regions to the apex in order to establish an infection in that location [11].

2.7.2 Differences in pathophysiology

Cavity formation is an important event in the TB life-cycle as the presence of cavities within the lung of the host greatly increases the chances of transmission to another person [250]. The general consensus within the

literature on cavitary TB is that a cavity arises when the caseous necrosis present in the centre of a TB granuloma (originating from the death of immune cells and destruction of surrounding tissue) enlarges and liquefies [35, 55–57, 106, 137, 250]. This liquefied caseum then empties out of the alveoli space, leaving behind an empty region. In this region, remaining bacteria are able to replicate in great numbers, possibly due to the sudden increase in availability of oxygen, as the cavity is in direct communication with the bronchial tree [96]. Bacteria are believed to be able replicate much better in liquefied caseum [96, 250], and coughing may distribute this bacteria-laden liquid caseum into other regions of the lung, or out of the body and into the air, to be inhaled by other people and transmit TB infection. However, there is evidence to suggest that the pathophysiology of post-primary TB differs significantly from primary TB. Whilst the caseating granuloma is a hallmark of primary TB, the classical paradigm that cavities are formed from a caseating granuloma eroding into an airway has been challenged [120–124, 159], due to evidence from the pre-antibiotic era where necropsies of post-primary TB patients displayed pneumonia [160].

Regardless of the actual events that occur during post-primary TB, the difference in localisation during primary and post-primary stages raises important questions relating to TB pathogenesis viewed over the whole of the lung:

- Assuming all bacteria land in the basal regions initially, how do bacteria disseminate across the lung from their initial location at the basal lung regions to establish post-primary infection at apical lung regions?

- Why is post-primary disease confined to the apex of the lung? What heterogeneous factors of anatomy occur within the lung to create a region at the apex that is preferential for TB proliferation?

These two questions are addressed in the following two Sections, 2.8 and 2.9, respectively.

2.8 Bacterial dissemination

Dissemination of *M. tuberculosis* bacteria can occur through a variety of methods [195], including:

- Local progression
- Lymphogenous progression
- Lymphohaematogenous progression
- Bronchogenic progression
- Haematogenous progression

Local progression occurs when TB disease expands directly through tissue [195]. This can occur through the creation of cavities or when neighbouring lesions expand and coalesce [38].

Primary TB infections often include a secondary infectious focus within the lymphatic system [15, 16]. Thus, it is assumed that bacteria are able to disseminate from the primary lung focus to the regional lymph nodes via the lymphatics [42]. The formation of the adaptive immune response necessary to contain pulmonary infections requires migration of immune cells from the primary focus in the lung to regional lymph nodes (Section 2.6.1). But doing so provides a route of dissemination for lung pathogens: as APCs migrate to the lymph nodes, they also traffic with them any pathogens that remain alive inside them [130].

The lymphatic system has multiple important functions within the human body, including the reclamation of blood plasma, blood filtration and initiating an adaptive immune response. It is a one-way system, with two terminal ducts: the right thoracic duct drains the right arm, upper right torso and head, while the thoracic duct drains the remainder of the body. These ducts empty into the right and left subclavian veins, respectively [66]. Doing so allows any immune cells produced by the lymphatic system to enter the bloodstream and reach their intended targets.

Therefore, by being transported to the lymphatic system, *M. tuberculosis* is provided with access to the bloodstream. The blood in the subclavian veins is deoxygenated, and is sent (via the heart) back into the lungs to be re-oxygenated before returning first to the heart and then to the remainder of the body. Any bacteria that pass through the entirety of the lymphatic system enter this blood and are sent back to the lungs, allowing for re-seeding in previously uninfected areas, or may infect the rest of the body. In the lung, lesions formed by this spread are termed 'Simon' foci [195].

Bronchogenic progression occurs when a TB lesion erodes into an airway, usually into a bronchus but also into the trachea. Soft caseous material expelled from the lesion is usually removed from the lungs as sputum by coughing, but some of this material may be aspirated and thus seed a new part of the lung. In this case, the new lesions that are formed are usually clustered together and may be of varying sizes [195].

Not to be confused with lymphohaematogeneous progression, haematogenous progression occurs when a lesion within the lung erodes directly into the bloodstream [195]. This allows for a larger number of bacteria to be disseminated through the blood to all areas of the body, and the term 'miliary' TB[1] is used to define all forms of haematogenous spreading of TB [163]. Cases where the bacteria establish infection in organs other than the lungs are defined as 'extrapulmonary' TB. This is not contagious, unless the person also has pulmonary infection or the disease has spread to the oral cavity [41].

2.9 Impact of physiology on TB

TB pathology occurs in different regions of the lung during different stages, and it is important to note that the physiology of the lung is not homogeneous within an individual. All human lungs display significant

[1]The term 'miliary' derives from the resemblance of the lesions to millet seeds

heterogeneity throughout, in the form of size, shape and function.

2.9.1 Ventilation

The main purpose of the lungs is oxygen exchange: air is breathed in and passes through the trachea and bronchi and eventually reaches the alveoli, where the oxygen in the air is exchanged for carbon dioxide released from the blood, which is then exhaled. The delivery of the air to the bronchi is heterogeneous: more air is delivered to the lower regions of the lung than to the apices, due to factors such as gravity [236] and asymmetrical branching anatomy [81] (see Figure 2.7, blue line).

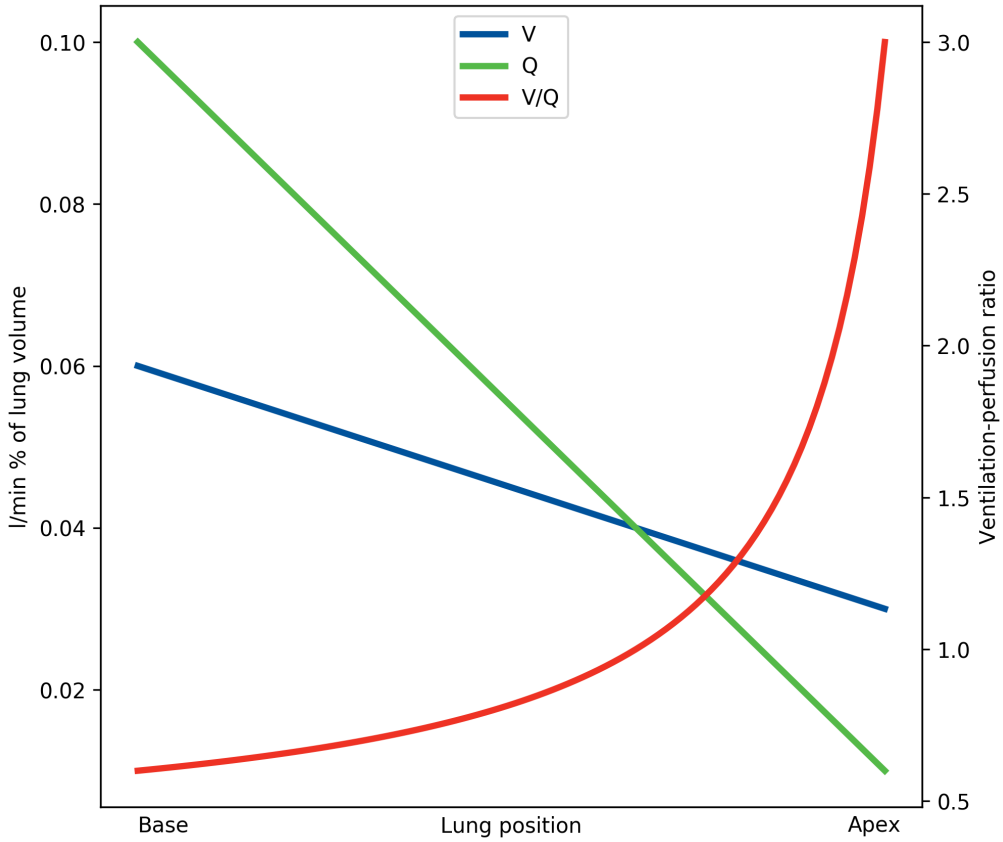


Figure 2.7: Distribution of ventilation (V) and perfusion (Q) in the lung, and differential in ventilation-perfusion ratio (V/Q). Both V and Q are higher at the base than the apex, but the differential in Q is much greater than the differential in V, resulting in a V/Q imbalance, such that V/Q is much higher at the apex than the base (taken from [237]).

Differentials in ventilation impact on TB pathology, particularly in its conception: a greater distribution of air towards the lower regions results in a greater deposition of inhaled particles at those regions. This explains the propensity for primary TB to occur towards the base of the lung, as this is the region where more of the aerosols containing *M. tuberculosis* bacteria land when breathed in.

2.9.2 Perfusion

Oxygen exchange requires de-oxygenated blood to be supplied to the lungs, and similarly to ventilation (Section 2.9.1), this supply of blood (often labelled as 'perfusion') also varies in the lung, with greater perfusion occurring at the lower regions than the apices. This occurs due to gravity [166], hydrostatic pressure differences [234] and anatomical effects [81] (see Figure 2.7, green line).

The rate of blood flowing to a region of the lung impacts on the body's ability to establish an efficient immune response there, as immune cells are recruited to sites of infection through the blood. This may explain, in some part, the apical localisation of post-primary TB: at the apices, the perfusion rate is at its lowest and thus the immune response there is expected to be lower than in other regions. Therefore, *M. tuberculosis* bacteria may favour the apex due to the lack of intervention from the immune system restricting bacterial growth.

2.9.3 Oxygen tension

Oxygen exchange within the body is imperfect due to the differences in scales between ventilation and perfusion (see Figure 2.7). Both ventilation and perfusion exhibit linear differential scales whereby they increase travelling down the lung, but the scale of increase is much higher for perfusion than for ventilation. Thus the apex of the lung has almost no perfusion, whilst the base of the lung is over-perfused. The result of this is that although more air is sent to the base of the lung, all oxygen within

the air there is transferred to the blood; whilst at the apex of the lung, the lack of perfusion means that some oxygen remains in the air before it is exhaled. Air that is inhaled but does not take part in oxygen exchange is termed 'alveolar dead space'. The physiology creates a differential in oxygen tension: the apex of the lung has a high oxygen tension whilst the basal regions have low values.

As noted in Section 2.3, *M. tuberculosis* is aerobic, and thus the high oxygen tension at the apex may make it a preferential environment for the bacteria, allowing them the oxygen they require to replicate excessively. The high oxygen tension at the apex is often cited as the cause of post-primary TB localisation at the apex, although the reality may be a combination of this and other factors.

2.9.4 Drainage

The drainage of immune cells from the lung tissue to the lymphatics is an important part of fighting an infection, both enabling the activation of an adaptive immune response, and clearing the lung tissue of pathogens and transporting them to the lymph nodes, where an abundance of lymphocytes allows easier destruction of the pathogen. During breathing, the lung expands and contracts, and this movement acts as a pump to push immune cells into the lymphatics. The movement is different throughout the lung, with the uppermost ribs moving the least and thus creating less drainage than at the lower regions [94].

This drainage differential could also impact TB progression, with the inability of the body to sufficiently drain cells at the apex possibly contributing to the accumulation of infection there.

Drug	Properties	Months
Isoniazid (H)	• Highly bactericidal against replicating bacteria	1-6
Rifampicin (R)	• Bactericidal against bacteria in both intracellular and extracellular locations • Greater sterilising potency against dormant bacteria than isoniazid	1-6
Pyrazinamide (Z)	• Potent sterilisation effect at high pH locations	1-2
Ethambutol (E)	• Works synergistically with other drugs to delay or prevent emergence of drug resistance	1-2

Table 2.1: First line drugs used for TB treatment, showing properties and their usage in particular months of the standard regimen [3, 244, 246, 249].

2.10 Treatment

2.10.1 Current treatment

Treatment of active TB typically requires multiple antibiotics over a sustained period of time. The WHO recommends a variety of regimens depending on the status of the patient. For patients with known drug-susceptible strains of the disease who have not received prior treatment, a standard regimen of 6 months is recommended, known as 2HRZE/4HR [246]. This requires an initial 2 month intensive period consisting of 4 drugs taken simultaneously on a daily basis, followed by a 4 month sterilising period of 2 drugs. The initial intensive phase is designed to destroy all bacteria that are in a replicating phenotype, whilst the sterilising phase eradicates any bacteria that are dormant (and thus more resistant). Table 2.1 lists all drugs used in the standard regimen, their individual purposes and their months they are prescribed for. Different regimens are recommended for other forms of disease, including latency [252] and patients with antibiotic resistance [84].

2.10.2 Antibiotic resistance

One of the major threats to global control of TB is the emergence of drug resistance [175]. This phenomenon is caused by mutation of the bacteria during cell division [89], and was discovered in *M. tuberculosis* shortly after the discovery of the first anti-TB drug, streptomycin [52]. Combination therapy (the use of more than one antibiotic in the treatment regimen) was subsequently shown to prevent resistance emerging [89].

Drug-resistant strains of TB are classified by the drugs that they are resistant to. Strains that are resistant to both isoniazid and rifampicin are classified as Multidrug-resistant tuberculosis (MDR-TB) [186]. The emergence of MDR-TB has serious implications, as these two drugs make up the continuation phase of the standard treatment regimen. Resistance to either one of these can be combatted by extending the standard regimen and using other standard first-line drugs (Table 2.1), but resistance to both requires the use of more toxic second-line drugs, such as fluoroquinolones [186]. Further resistance to second-line drugs has even greater consequences. Extensively drug-resistant tuberculosis (XDR-TB) is classified as resistance of the bacteria to at least isoniazid, rifampicin, any fluoroquinolone, and at least one of three injectable second-line drugs [126].

2.10.3 New treatments

Delivering a clinical trial of a new treatment, for any disease, is a long and complicated process with no guarantee of success. The trials take a long time to complete, involving lengthy time periods between treatment and follow-up [184]. This is particularly the case in TB, where the slow-growing and resistant nature of the *M. tuberculosis* bacillus requires a lengthy treatment phase. Furthermore, in order to make any significant conclusions trials require large sample sizes of patients. The total cost of running a clinical trial can be very high, reaching up to USD$600 million [70]. One form of clinical trial that is gaining more momentum with TB research is non-inferiority trials. These typically involve a trade-off, a new regimen is trialled which, whilst perhaps less efficacious than

the current standard, has other benefits, perhaps in duration or less side-effects [220]. A new regimen can be implemented if it is shown to be as effective (i.e. not inferior) or better than the current standard regimen, based on some pre-defined metric and margin of non-inferiority.

The standard regimen for TB requires patients to take daily doses of chemotherapeutic drugs for a period of 6 months. This can be difficult for patients to adhere to [5, 253], and non-adherence can have serious complications, both for the patient (due to relapse) and the wider community (due to the risk of antibiotic resistance emerging). Thus shortening the duration of treatment regimens is seen as a priority, as a regimen of, for example, 4 months is expected to improve patient adherence and lower costs [90]. To this end, regimens utilising drugs outside of the current gold standard have been trialled, aiming to show a shorter regimen is non-inferior to the current standard. Moxifloxacin is a drug with promising bactericidal activity against *M. tuberculosis* in *in vitro* studies, and this drug was used in clinical trials (REMoxTB) to test whether utilising it in place of one of the current standard drugs could reduce treatment time to 4 months. The trials were unable to prove non-inferiority [90], possibly due to the inability of the drug to penetrate into caseum within lesions [196].

The drug development process for TB is thus long and complex and in dire need of streamlining [180]. Whilst changes could be made to improve the quality of non-inferiority test end-points and margins [184], these changes will only be able to improve the quality of results of regimens put to trial: they will not be able to mitigate against the risk that the regimen will not be successful. One way to make improvements in the length of the drug development process would be to improve the prioritisation mechanisms by which regimens are chosen.

Consider the drug development process shown in Figure 2.8. A novel treatment regimen is conceptualised, and then experiments (typically *in vitro* lab experiments or *in vivo* animal experiments) can be used to determine the likelihood of success. The regimen can be put to trial, and

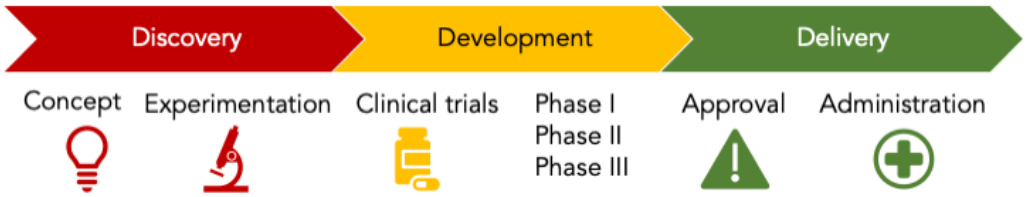


Figure 2.8: The drug development process

if the trial is successful, approval can be granted and the drug can be administered to patients globally. One of the key bottlenecks here is the experimentation stage - the lack of a suitable animal model that accurately reflects human pathophysiology hinders the predictive capability at this stage. Therefore, improvements at the experimental stage could lead to better predictions and thus reduce the risk that an expensive development stage producing a negative result.

2.11 Conclusion

In this chapter, we have outlined some of the major elements concerning TB, including the history, transmission, immunology, pathology and treatment of the disease. In particular, we have explained how different stages of the disease have distinct localisations within the lung, and have outlined the biological factors which are believed to contribute to these. We have explained how the standard treatment regimen is long and difficult to adhere to, and how efforts to reduce this treatment time have been unsuccessful. We have described the inherent costs in terms of time and money associated with clinical trials, and have shown how improvement in the discovery stage with regard to make predictions over the likelihood of efficacy of a new regimen could reduce these costs. In the next chapter, we look at how *in silico* models could contribute in terms of improving predictions at the discovery stage.

Chapter Three

Networked metapopulation framework

In which we provide an overview of concepts relating to metapopulation and mechanistic modelling, present MetapopPy, a framework for creating networked metapopulation models with heterogeneous environments, and detail our verification testing of it against published models.

3.1 Introduction

As discussed in the previous chapter, TB pathology within the human lung involves a complex series of dynamics and interactions. The environmental heterogeneity present within the lung is believed to impact on the distinct localisations seen during different stages of infection, but the exact contributions it has on disease progression are poorly understood. Mechanistic modelling of real-world process can allow us to explore the dynamics of a system, helping us to shed light on these complexities and improve our knowledge, thus guiding us in our decisions regarding the direction and scope of future research.

In this chapter, we provide an overview of the current state of *in silico*

models of within-host TB dynamics, showing that most models to date have focused on small scales, modelling single lesions. Those that have taken a larger view of disease over the whole lung have not yet incorporated the environmental heterogeneity that is believed to be a crucial part of the disease pathology. We show how models can be built incorporating environmental heterogeneity, creating a metapopulation of spatially distinct sub-populations, with the local environment of each patch being assigned different environmental attributes, thus creating differentials of resources across the landscape. We overview some modelling processes that are relevant to this thesis, outlining the process for simulating time using the Gillespie algorithm.

We then present a framework for creating a networked metapopulation, written in Python, which allows a population to spread across distinct patches, each of which may be assigned its own values for environmental attributes. These attributes may then affect the event dynamics, modelled using the Gillespie algorithm, that occur across the network of patches and edges, where members interact within a patch or translocate between patches. Thus the heterogeneity of the environment affects the dynamics occurring upon it, an important factor when modelling TB across the lung. Finally, we outline verification testing performed on the model against existing epidemic models at both a single-patch and multiple-patch scale to demonstrate the validity of *MetapopPy* as a means of simulating dynamics upon a heterogeneous environment.

3.2 Background

3.2.1 *In silico* models of within-host TB

Mathematical and computational models within TB research have focused on three main aspects of the disease [132]:

- The epidemiology of the disease, studying the spread of TB over populations of people [229].

- The bacteriology of the disease, particularly studying the growth, drug-resistance and latency phenomena of the *M. tuberculosis* bacteria.

- The pathology, immunology and treatment of the disease, exploring the complex interactions and dynamics that occur within the host during infection.

In this section, we review the latter, detailing some of the *in silico* models of TB within-host dynamics (see [132] and [36] for comprehensive reviews on this subject).

3.2.1.1 Lesional-scale models

The first models exploring the evolution of a single lesion within the lung were agent-based models (ABMs), in which the cells and bacteria which constitute the system are represented as individual components with their own rules ('agents') and a lesion forms as an emergent behaviour from the combination of the interactions of these agents [200, 213]. These models were able to study the formation of a granuloma by taking into account the spatial factors, such as overcrowding mechanics and the spatial location of T cells [213] which impact the disease dynamics, which had been impossible to do through mathematical differential equation models. Other methods such as partial differential equations (PDEs) [73] and metapopulations [82] have also been used to study lesion progression.

Segovia-Juarez et al [213], described a model (later named *GranSim*) which would form the basis of many future models of TB lesions. The environment modelled is a 2D portion of alveolar tissue, 2mm by 2mm in size, divided into a lattice of micro-compartments. Each compartment may contain a 'source' (whereby immune cells enter the system), extracellular bacteria or immune cells. Cells are able to move and may come into contact with bacteria, and thus ingest them. The bacteria then become intracellular and may replicate within the immune cell. Whilst this model is an ABM, it also incorporates some elements of cellular automata. A chemokine gradient occurs whereby immune cells secrete

chemokines and these diffuse between micro-compartments. This is achieved through a cellular-automaton like process, where the chemokine value at a micro-compartment changes over time based on the values in the micro-compartments adjacent to it. Time is modelled synchronously, with each time-step possibly involving multiple events.

Further models extended the initial *GranSim* model by creating a multi-scale model examining dynamics at tissue, cellular and molecular scales by including ordinary differential equations (ODEs) and PDEs to model cytokine molecules [72]. Inclusion of pharmacokinetic (PK) and pharmacodynamic (PD) effects into the *GranSim* model showed that antibiotics can reach insufficient quantities within the granuloma between dosages, with antibiotic gradients occurring within the lesion [190].

Models by other groups were produced, including those exploring the effects of oxygen and nutrient starvation, whereby the structure of the granuloma contributes to the bacteria being deprived of access to these important resources [59, 214]. ABM models have also been used to study the effects of spatial distribution of oxygen upon bacterial cell state [26]. This model showed that regions of the tissue may be preferable for bacterial proliferation, as they are located far enough away from important blood vessels that an immune response cannot be triggered fast enough and chemotherapeutic concentrations are too weak to destroy the bacteria. Importantly, this model also incorporates the notion of 'dormant' bacteria, with bacteria becoming dormant due to oxygen differentials within the lesional environment. These findings indicate the importance of spatial structure to the progression of a TB lesion.

3.2.1.2 Inclusion of the lymphatics

Primary infection with TB often includes infection of the lymph nodes, and transfer of immune cells (and possibly bacteria) between the lung and the lymphatics is an important part of the disease process for TB. As such, models have been created which include this translocation of cells and thus create a two-organ model. A two-compartment mathematical

model [151, 153] depicts the transfer of immune cells between the lung and lymph nodes. In the lung compartment, differential equations describe the growth of bacteria and their interactions with immune cells, and DCs are permitted to transfer to the lymph compartment, whereby they interact with naïve T cells and cause T cell activation. The activated T cells may then move to the lung compartment and influence dynamics there. This model demonstrated the importance of transfer of both DCs and T cells between these sites. A further model constructed from this two-compartment model also included bacterial infection within the lymphatics [154, 174] and macrophage infection there, showing how cytokines balance the macrophage phenotypes present in each environment.

In [152], the *GranSim* model was combined with the ODE model that described the priming of T cells by cells migrating out of the lung. Thus, the ABM of the lung tissue simulates the initial infection and the cell actions, with the number of infected cells serving as a proxy to determine the number of APCs that will be present in the lymphatics and this provides appropriate values to the equations. This model showed that the spectrum of infection outcomes (active, latency, clearance, etc.) can be regulated by key immune cell migration rates.

3.2.1.3 Modelling infection over the whole lung

Whilst the models described in section 3.2.1.1 explore a single lesion in the lung, there exist models that have explored TB development over a larger, whole-organ scale. Most of the whole-lung models of within-host TB dynamics are ODE and PDE models, exploring *M. tuberculosis* densities during infection [10, 148], the factors which regulate the immune system [147, 219, 240] and the effects of co-infection of TB with HIV [17, 131].

Whilst these models are scaled to explore dynamics, it is important to note that none of these contain any spatial dimensions within the lung. In each, the lung is treated as a single entity: a homogeneous, well-mixed environment and thus the location where infection occurs is irrelevant. For example, in [148], the effects of environmental factors such as oxygen

and nutrient starvation on bacterial states (active, dormant and latent) over time is modelled to create states of active, latent and re-activation TB. But there is no indication that the actual environment where these disease states are occurring may be different, as described in the previous chapter.

3.2.2 Metapopulations

Many early population models, such as those used in ecology modelling, made an assumption that the entire population was well-mixed, with all members of the system able to interact with each other and the entire environment assumed to be homogeneous in nature. Whilst this is appropriate in some scenarios, in scenarios where a heterogeneous environment exists (such as the internal conditions within the lung as explained in Chapter 2) this paradigm can be too restrictive. In order to allow the differential spatial attributes of the environment to impact the simulation dynamics that occur, a paradigm which incorporates space is needed.

One such paradigm is that of metapopulations. A metapopulation is a population in which the members are divided based on their spatial location. In the original Levins model [139], the environment is divided into spatially distinct sites, commonly referred to as 'patches'. Each patch is either inhabited by a population ('occupied') or the population is extinct at said patch ('unoccupied') (see Figure 3.1A). Thus, local dynamics are ignored: if a patch is inhabited, it is assumed that the carrying capacity of the patch is reached. All patches contain the same extinction probability, and an unoccupied patch can become inhabited by migration from an occupied patch [105][1]. The Levins model can be used to predict the threshold at which a population will survive or become extinct.

The Levins model incorporates space in as much as sub-populations are spatially distinct and cannot interact with one another, but it does not include the environmental heterogeneity found in many real-world sys-

[1]These dynamics can be seen as analagous to a networked SIS epidemic model, with migration reflecting infection and extinction reflecting recovery.

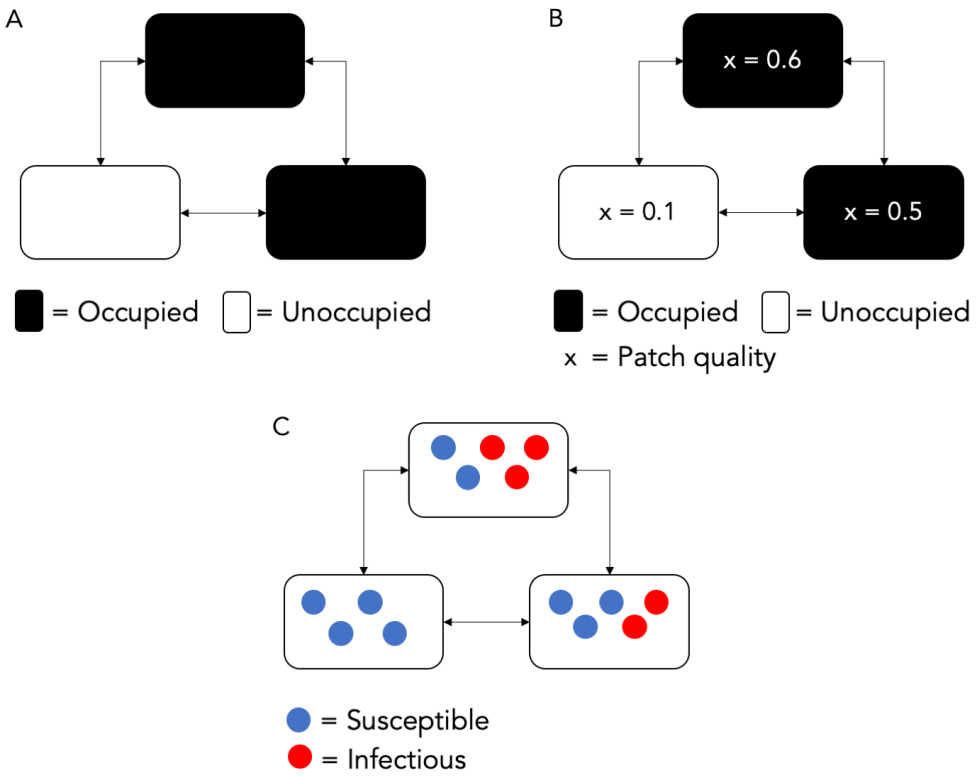


Figure 3.1: Examples of different types of metapopulation models (each with 3 patches). A) Levins model of occupied/unoccupied patches. B) Gyllenberg model whereby patches are assigned a quality value. C) Epidemic SIS model of Susceptible and Infectious members.

tems, as each patch contains the exact same dynamics as all other patches. Later metapopulation models incorporated heterogeneous environments by including patch 'quality' - the better the quality of a patch, the less likely the species is to become locally extinct there (see Figure 3.1B) [101].

The models described thus far have examined a single species, but metapopulation modelling can be used to study systems whereby multiple species interact with one another. These are of particular use in epidemiology, with each species (or 'compartment') representing a disease state [247]. Examples of these include Susceptible-Infectious-Susceptible (SIS) (shown in Figure 3.1C), Susceptible-Infectious-Recovered (SIR) and Susceptible-Exposed-Infectious-Recovered (SEIR) models, where each compartment represents a recognised state of the given disease, and

members pass between compartments as their disease progresses. Rather than a single set of equations defining the rate of change of disease states of the whole population, each patch may contain its own set of equations, and movement between patches (hereby termed 'translocation') is included within these equations, either impacting the chances of infection (due to short term migration) [143], or by the movement of members between patches (due to long-term or permanent migration) [46].

3.2.3 Time modelling

All simulation of real-world event systems requires us to consider temporal factors - at what point and what order do events occur? There are two main forms of event modelling: synchronous and asynchronous modelling [39, 207, 212].

3.2.3.1 Synchronous

In synchronous modelling, the simulated time is considered discrete and is broken up into slots, with all events in one time slot considered to occur at the same time. One such implementation of a synchronous time modelling paradigm is shown in Figure 3.2. Here, all events chosen in the time-step occur at the exact same time, i.e. between t and $t + \tau$.

3.2.3.2 Asynchronous

In asynchronous modelling, events are performed one-at-a-time, with no overlap between when events occur. In this case, the time difference between event occurrences can be infinitesimally small.

The Gillespie algorithm, sometimes known as the Stochastic Simulation algorithm, provides a means to model events asynchronously. In its initial description, the algorithm is used to describe molecular species interacting through chemical reactions [87, 88]. However, the principles can apply to any system where interactions occur between sub-populations, and this has been used in a variety of scenarios, particularly in epidemic modelling [97, 228], with the species representing disease states. For

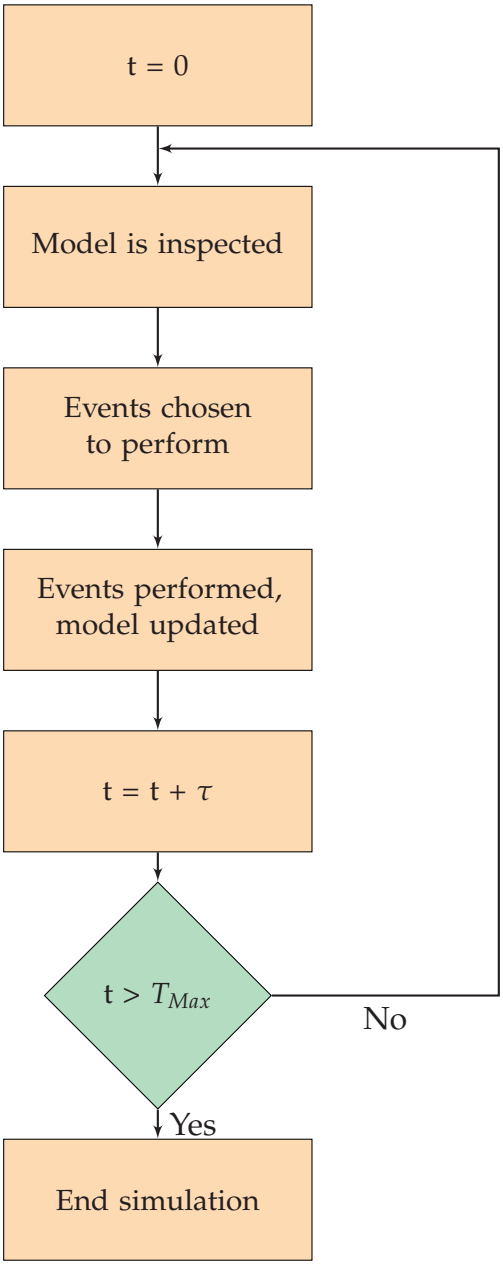


Figure 3.2: Example of synchronous time modelling. A time-step, τ, is determined before simulation as is the maximum time T_{Max}. Simulation begins with t=0, and an iterative loop continues whereby the state of the model determines all events to occur in the time-step. Simulation ends when the maximum time is exceeded.

the purposes of describing the algorithm in this thesis, we use a more generalised description: exploring the algorithm through the lens of *elements* of the system interacting through *events*. Here, we describe the algorithm for a system occurring in a single spatial location (description adapted from [32, 87, 88]):

Consider the set of N elements $\{S_1, S_2, \ldots, S_N\}$. These elements may participate in a series of M events, $\{E_1, E_2, \ldots, E_M\}$. The state of the system at time t is described by the vector $\mathbf{X}(t) = \{X_1(t), X_2(t), \ldots, X_N(t)\}$, where $X_i(t)$ is the value of element S_i in the system at time t. It is assumed that the system is well-mixed (i.e. all elements of the system can interact with all other elements)[2].

Each event, E_j, must define the following:

i) A *reaction parameter*, c_j, which gives the average probability that one E_j event will occur in the next infinitesimal time interval $[t, t + dt)$.

ii) A *state variable* function, $h_j(\mathbf{X}(t))$, which takes the current state of the system and determines the possible number of occurrences of event E_j.

iii) A *state change* vector, $\boldsymbol{\nu}_j = \{v_1, v_2, \ldots, v_N\}$, where v_i describes the changes that occur to the value of element S_i when event E_j is performed.

Using these, each event can calculate its 'rate', $r_j(t)$, at time t as per equation 3.1, which can be defined as the probability, to first order in τ, that an E_j event will occur in the next time interval τ.

$$r_j(t) = h_j(\mathbf{X}(t)) * c_j \tag{3.1}$$

[2]It is important to note that the term 'element' here is left deliberately vague: events may involve interaction between a variety of different objects within the system. One conceptual view, which we utilise later in the *MetapopPy* framework, is that the modelled environment contains a subpopulation of species, as well as attributes defining the nature of the environment in the system at the location being modelled. Both of these may impact on the rate at which events occur. Thus, the elements of the system are a set of both discrete, non-negative integer counts of members of various species and continuous positive or negative values for environmental attributes.

An iteration similar to that described in 3.2.3.1 occurs, but instead of incrementing by the same time-step iteration, a time-step, $\tau(t)$, is calculated dynamically based on the sum of all rates, $R(t)$, at time t, calculated in equation 3.2.

$$R(t) = \sum_{j=1}^{M} r_j(t) \tag{3.2}$$

This value is then used to calculate $\tau(t)$ as per equation 3.3, where q is a pseudo-random number generated for each simulation timestep in the interval [0,1) (see [87] (Appendix: The Inversion Generating Method) for the derivation of this formula from the appropriate probability density function).

$$\tau(t) = \frac{1}{R(t)} * ln\left(\frac{1}{q}\right) \tag{3.3}$$

The iteration for the Gillespie Algorithm proceeds as follows:

1. Simulated time starts at zero, i.e. $t = 0$
2. All events calculate their rates as per equation 3.1, giving a rate vector $\{r_1(t), r_2(t), \ldots, r_M(t)\}$.
3. $R(t)$ is calculated as per equation 3.2 and $\tau(t)$ is likewise calculated as per equation 3.3.
4. An event, E_j, is chosen probabilistically based on values of the rate vector, with a greater rate indicating a greater likelihood of being chosen.
5. The state change vector, $\boldsymbol{\nu}_j$ associated with E_j is used to update the state of the system.
6. Time is incremented, such that $t = t + \tau(t)$.
7. Steps 2 to 6 repeat until a time limit is reached.

3.2.3.3 Comparison of synchronous and asynchronous methods

Whilst synchronous and asynchronous methods can be used to model the same scenarios, the choice of time modelling paradigm can lead to significantly different dynamics [212], and the choice of updating

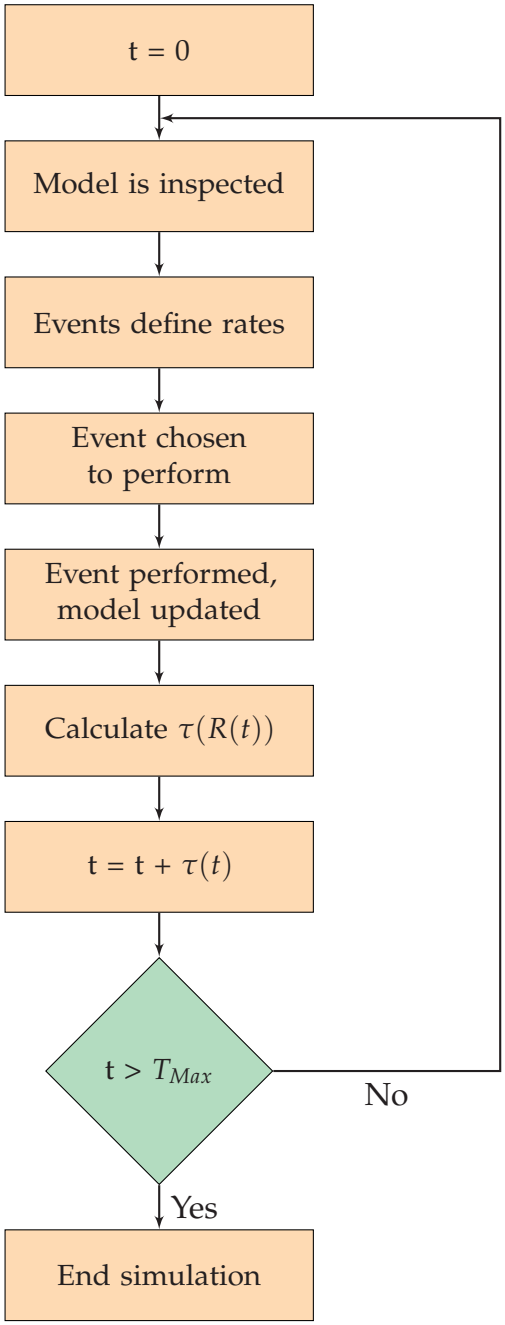


Figure 3.3: Example of asynchronous time modelling. Contrasting with synchronous time-modelling, only a single event is performed per loop, and the time-step, τ, is now a function of the event rates $R(t)$ and is determined each loop.

method should depend on the real-world system being modelled [39]. Asynchronous methods are efficient in sparse situations: when not many interactions occur, no time is wasted by simulating time-steps where no events happen [228]. Furthermore, asynchronous modelling avoids any issues around conflicting events - as events occur one-by-one, there is no scenario where two mutually-exclusive events could both occur (as may happen in a synchronous system), which would lead to a paradoxical outcome, and it has been argued that asynchronous methods present a more realistic approximation of real continuous time [39, 49, 212] and we therefore chose to use an asynchronous updating method within our modelling framework.

3.3 Methods

In this section, we detail the creation of a networked metapopulation framework, name *MetapopPy*. This framework, written as a Python library, is abstract in nature and allows for the creation of a metapopulation consisting of patches connected together with edges, which signify translocation of member between patches. Event dynamics can be created, and a Gillespie Algorithm implementation is used to simulate these dynamics occurring over the network. In order to create this heterogeneous environment, we chose a networked metapopulation paradigm. As shown in section 3.2.2, using a metapopulation allows for the spatial discretisation of the entire landscape, creating distinct regions, each of which can be positioned within the larger environment and given environmental attributes that are appropriate for that position. Whilst metapopulations have been used extensively for modelling populations of people, we used this paradigm to model cellular-level interactions within the lung during TB infection, a process which has only been done rarely [82] and never at the scale of a whole lung including the environmental heterogeneity discussed above.

3.3.1 Existing software

The *MetapopPy* Python library builds upon already existing Python libraries. Firstly, *epyc* [63] serves as the foundation upon which *MetapopPy* is built. *epyc* is a Python module designed to automate the control of long-running experiments with multiple repetitions. This module separates the logic of a single experiment from the processes required to run multiple simulations of that experiment with varying parameters, which proved to be ideal for our own framework. As such, the base `Dynamics` class of *MetapopPy* is a subclass of the `Experiment` class of *epyc*, and so can be plugged efficiently into *epyc* experiment combinators to allow various parameter spaces to be explored, allowing us to focus on the development of creating a metapopulation framework without the concern of how to implement modules to run simulations.

In *epyc*, the parameter space can be explored as follows:

1. An instance of the `Lab` class is created and assigned a set of parameters, with each parameter having either a single value or a range of values.

2. A parameter space is created from the parameter values, with the default being the cross-product of all possible parameter values.

3. A parameter sample is taken from the space. This is then passed to a given experiment (an instance of the `Experiment` class).

4. The experiment calls its `configure` method to prepare for the simulation. This occurs once for every parameter sample.

5. For each parameter sample, a number of experiment repetitions may be run (through the `RepeatedExperiment` class). For each repetition, the `setUp` method the experiment is called. This may be used to initialise the experiment for the given repetition.

6. The actual mechanics of the `Experiment` instance occur using the `do` method, and results are recorded.

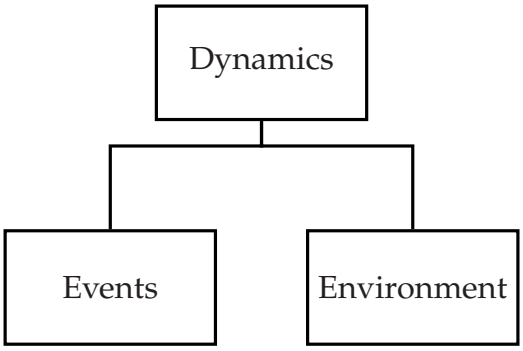


Figure 3.4: Overview of the modules within the *MetapopPy* framework

Thus, the `configure` and `setUp` methods are used to initialise the model (for a parameter sample and a repetition, respectively), and the `run` function performs the actual experiment. These functions are used within the *MetapopPy* framework, as described in Section 3.3.2.3.

The metapopulation network of our framework is created by using the *NetworkX* library [102], which is used to create and manipulate complex network structures of nodes and edges. Subclassing the `Graph` class of *NetworkX* provided us with a network with useful manipulation interfaces with which to add and remove nodes and edges.

3.3.2 Framework description

MetapopPy consists of three main modules (see Figure 3.4):

- `Environment` - the environment upon which the populations being modelled reside, which will be altered by the events
- `Events` - the individual actions and occurrences that alter the state of the system
- `Dynamics` - the combination of the events and the environment

In this section, we briefly overview each module to explain how the framework is used.

3.3.2.1 Environment

The `Environment` class of *MetapopPy* is a subclass of the `Graph` class of *NetworkX*, and thus takes the form of a network of patches and edges[3] which may be altered dynamically. At initialisation, the following must be provided to the `Environment` class:

i) A list of population *compartments*, reflecting the divisions within the overall population (by species, status etc.).

ii) A list of *patch attributes*, reflecting the environmental conditions at a patch.

iii) A list of *edge attributes*, reflecting the environmental conditions at an edge.

When a simulation is started, the environment automatically assigns dictionaries of compartments and patch attributes to each patch of the network, and a dictionary of edge attributes to each edge. The default position is that all patches contain the same compartments within their populations and the same types of patch attributes, and all edges contain the same types of edge attributes. We provide a subclass, `TypedEnvironment`, that assigns a 'patch type' to each patch, in order to make an easier distinction between different parts of the environment (e.g. between different organs within the body), and each patch of a given type is assigned only the patch attributes that are relevant to that type.

3.3.2.2 Events

Following the Gillespie algorithm (see section 3.2.3.2), each subclass of the abstract Event class of *MetapopPy* must:

i) provide a *reaction parameter* value when declared. This value represents the probability of a single instance of the event occurring in the given unit time. For the initial framework, this was a static value fixed at the start of simulation.

[3] *NetworkX* describes its network in terms of 'nodes' and 'edges', as is common in the Complex Network literature. Metapopulation literature typically describes the environment in terms of 'patches' and 'edges', and we choose to use this nomenclature as the framework is concerned with creating metapopulations.

ii) provide a function for calculating the *state variable*. This function will take the current state of a single patch of the network and return a float value, representing the number of possible occurrences of the event at the current time. The *rate* is automatically calculated by multiplying the reaction parameter by the state variable.

iii) provide a *perform* function. This can be seen as analagous to the state change vector described in 3.2.3.2, and is a function that updates the network when given a patch on which the event is to be performed. It should be noted that events may update more than one patch, as in the case of translocation of members from one patch to another.

3.3.2.3 Dynamics

The `Dynamics` class is a subclass of the 'Experiment' class of the *epyc* library. Therefore, this class forms the base of a single experiment - a range of values can be provided for parameters, and the parameter space can be explored by performing simulations using different combinations of parameter values. The `Dynamics` class of *MetapopPy* requires the following:

i) an instance of the *MetapopPy* `Environment` class. This instance may already have a network structure, or otherwise may create the network structure dynamically at simulation execution-time based on the parameter values assigned.

ii) a function to create the events that will occur over the network. These events are then stored within the `Dynamics` instance.

iii) functions to determine the initial conditions of the network. These are expected to take the parameter values being explored and return a dictionary of the appropriate seedings of compartments and attributes at the beginning of the simulation.

The *MetapopPy* `Dynamics` class overrides the methods present in the *epyc* `Experiment` class to perform its necessary functionality:

- At the `configure` stage (i.e. when a new parameter sample has been selected), the `Dynamics` class checks to see if a network has

been provided. If not, then one is created (using a user-defined process and the given parameter sample). Update handlers are attached to the network. These are lambda functions[4] and will fire if a change is made to the network, informing the `Dynamics` instance that a change has occurred and rate values must be recalculated. An initial seeding for all patches and edges is calculated based on the parameter sample (using a user-defined process).

- At the `setUp` stage (i.e. at the start of each of the repetitions for a parameter sample), the network is reset to an empty state (as it may have been altered by previous repetitions) and then updated with values from the seeding.

- At the `do` stage (i.e. the actual repetition itself), the start time is set, as is the maximum time that will be simulated. A rate table is created, with rows representing the patches of the network, and columns representing the events. Thus, every patch-row combination has a value, the rate of the event at that patch.[5] The following loop iteration then occurs repeatedly until the simulated time exceeds the maximum:

    1. A time-step value, τ, is calculated using the sum of all values in this rate table, following equation 3.3.
    2. An event and patch combination is chosen probabilistically using the individual rate values in the table.
    3. The `perform` function of the event is called using the patch as an argument. The event then performs the necessary actions. Where the event changes values on the network, this triggers the update handler functions which returns a list of the patch(es) and compartment(s) and attribute(s) that have

[4] A lambda function in the context of the Python programming language is an anonymous function that can be passed and called by another function. These are used within *MetapopPy* as they allow the `Environment` object to update the `Dynamics` object without having to see its internal functions and/or structure.

[5] For the case of translocation events moving members between patches, we assume that the event occurs at the patch where the member originally resides

been changed to the dynamics. The dynamics then updates the rate table for events occurring at the altered patches.

4. Time is incremented by τ.

3.3.2.4 Mechanics to improve efficiency

Whilst the above is sufficient to run experiments, the following functionality was also added to improve performance:

- Patches may be considered 'active' or 'inactive'. Only when a patch becomes 'active' is it added to the rate table. When 'inactive', no event dynamics may occur at a patch. This can be used to focus simulation only on the areas of interest in the network. As such, a patch is only seeded with values when it becomes active. The default option is that all patches are always considered 'active', but this may be overridden with user-defined functions.

- In order to reduce computational overhead when updating the rate table, each event must list the compartments, patch attributes and edge attributes it is dependent on to calculate its state variable. Then, when an update occurs, only the events which are dependent on the changed items are called on to recalculate their state variable (and thus also recalculate their rate).

- A time-step is defined by the user as to when to record the state of the network. Once the simulated time exceeds a multiple of this time-step, the state of the network is recorded to the results. This avoids recording excessive amounts of data where only minimal changes have occurred.

- In the case where a `TypedEnvironment` instance has been used, the patch type may be used to drive the model dynamics, such that events can be restricted to occurring only at a given patch type. This is done with a `PatchTypeEvent` class, which is initialised with a patch type, and returns a zero value for its state variable function if the supplied patch is not of the same patch type.

3.4 Verification testing

In order to ensure that the *MetapopPy* framework was suitable for modelling dynamics over heterogeneous environments, we examined the literature on epidemics to find models which could be used for comparison tests. We chose this field as metapopulation paradigms have been used extensively to simulate epidemic spreading (as opposed to, for example, within-host disease progression, for which few models exist). We used the *MetapopPy* framework to recreate the models within these sources and verify that the framework creates suitably accurate results compared to deterministic mathematical models. In this section, we review models at single-patch and multi-patch scales and compare the results from those sources against our own reproductions. We also describe the construction of the *MetapopPy* model used in each instance, in order to demonstrate how a conceptual model can be turned into the necessary class instances within *MetapopPy*.

3.4.1 Single patch epidemics

Most early epidemic models are essentially single-patch models: the entire population is constrained to a single location and it is assumed that the population is well-mixed. These models have been used extensively for TB modelling, with the dynamics often presented as a form of SEIR model, with the Exposed compartment representing the latent stage of the disease.

For the first verification test, we used *MetapopPy* to recreate a single patch infection model of TB, as described by Herrera et al [110]. This model was created to investigate TB dynamics in semi-closed communities (communities with some departure and recruitment but which have the potential for sustained daily contact between members over periods of weeks, months and years). This model contains 5 compartments: Susceptible (S), Exposed (E), Infected and infectious (I_I), Infected and Non-infectious (I_N) and Recovered (R). Firstly, we recreated the graph present in the original paper (Fig 5 of [110]) using the differential equa-

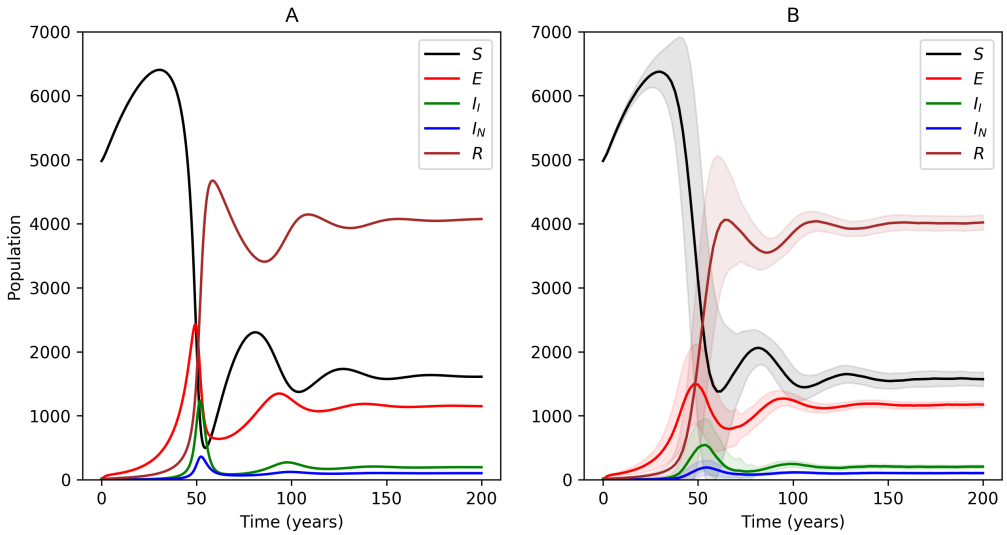


Figure 3.5: Side-by-side comparison of results of a single-patch epidemic model created by A) differential equations and B) *MetapopPy*. The equations for the model in A were taken from [110]. These were converted into the necessary events for a *MetapopPy* model. B) shows results from 200 *MetapopPy* simulations: solid line is mean value, shading is error.

tions described by the authors, the results of which can be seen in Figure 3.5A.

We then converted the equations of the Herrera model into Events within a *MetapopPy* model. Figure 3.6 shows the transition between compartments within the original Herrera model. Each arrow, representing a member leaving and/or entering a compartment, can be seen as a unique event, thus needing an instance of the `Event` class in *MetapopPy*.

Any event that moves members from one compartment to another would require two corresponding terms in the equations of an equation-based model (a term for the compartment the member is originally in to remove the member, and a term in the new compartment to add the member). However, in *MetapopPy*, this type of event only requires a single event, defining the rate at which the event occurs and the impact it has on the system. As an example, consider the infection events of the Herrera model. S members come into contact with I_I members and become

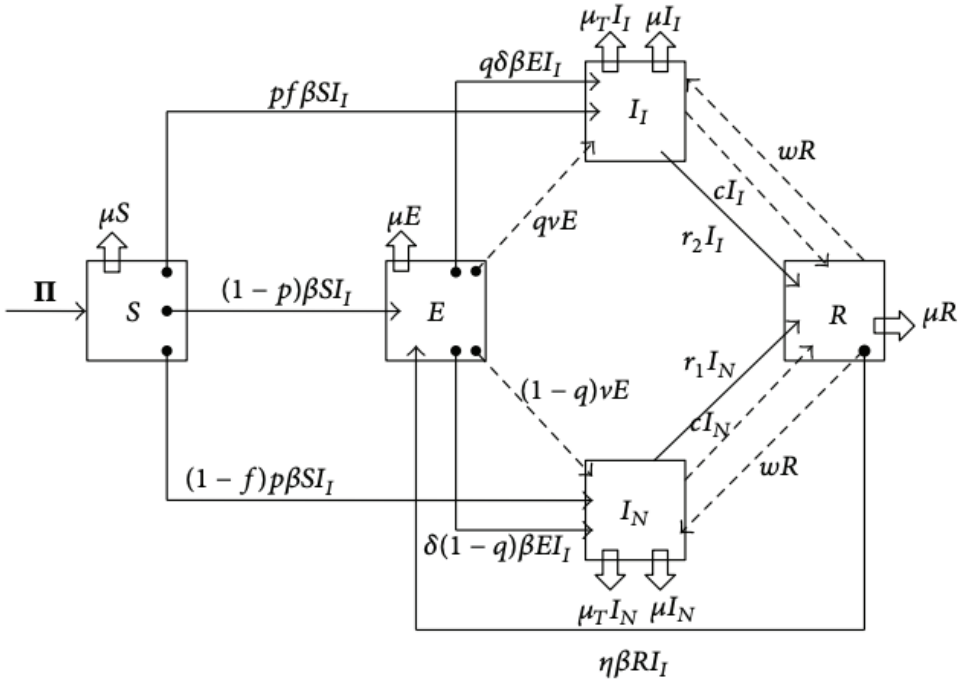


Figure 3.6: Flowchart of the movement of members between compartments in the Herrera model (taken from [110]).

infected. Of these newly infected members, a proportion, p, become actively infected, whilst the remainder, $1-p$, become latently infected (E). Of those actively infected, a proportion, f, become infectious and can transmit disease (I_I), and the remainder $(1-f)$ become non-infectious (I_N). Thus 4 terms are needed:

- $-\beta SI_I$ in $\frac{dS}{dt}$ to describe members becoming infected and leaving the S compartment
- $(1-p)\beta SI_I$ in $\frac{dE}{dt}$ to describe members becoming latently infected and joining the E compartment
- $pf\beta SI_I$ in $\frac{dI_I}{dt}$ to describe members becoming actively infected and infectious, joining the I_I compartment
- $p(1-f)\beta SI_I$ in $\frac{dI_N}{dt}$ to describe members becoming actively infected but non- infectious, joining the I_N compartment

This can be implemented in at least 2 ways within *MetapopPy*:

1. A single event which occurs at rate $\beta S I_I$. This represents all infection routes. Then, the perform function can be set to probabilistically choose a compartment to place the newly-infected member into, with probabilities of $(1-p)$, pf and $(1-f)p$ for E, I_I and I_N respectively.

2. Three separate events, one each for conversion to E, I_I and I_N. These will occur at rates $(1-p)\beta S I_I$, $pf\beta S I_I$ and $p(1-f)\beta S I_I$ respectively, and thus the reaction parameters set to use the appropriate parameters for each (this method was chosen for our model).

Having created the events, we created an *Environment* instance which consisted of a single patch which we seeded with the appropriate values (4980 S members and 20 I members) at the beginning of simulation. 200 simulations were run, and the averaged results with standard deviation can be seen in Figure 3.5B. As shown, *MetapopPy* provides a reasonable approximation of the same model created using differential equations.

3.4.2 Multi-patch epidemics

In order to verify that *MetapopPy* works as an approximation for modelling metapopulations, we proceeded to compare results against a model that included multiple patches in a heterogeneous environment. In [157], the authors define a model for studying a SIR disease occurring over multiple patches, each of which has different parameter values for events such as infection and recovery[6] and with migration of members between each patch. The equations that compose this model are similar to those of a single-patch model, but each include terms to alter the subpopulations based on the counts of same-compartment members in other patches. In this model, migration is driven by the number of connections on the network - the more edges a patch has, the greater the rate of migration of members out of the patch (and similarly, the more influx of members from other patches.) Figure 3.7A, C and E show the results taken from

[6] In an epidemic context, variable rates could occur for a variety of reasons, such as due to regional/national differences in healthcare or poverty.

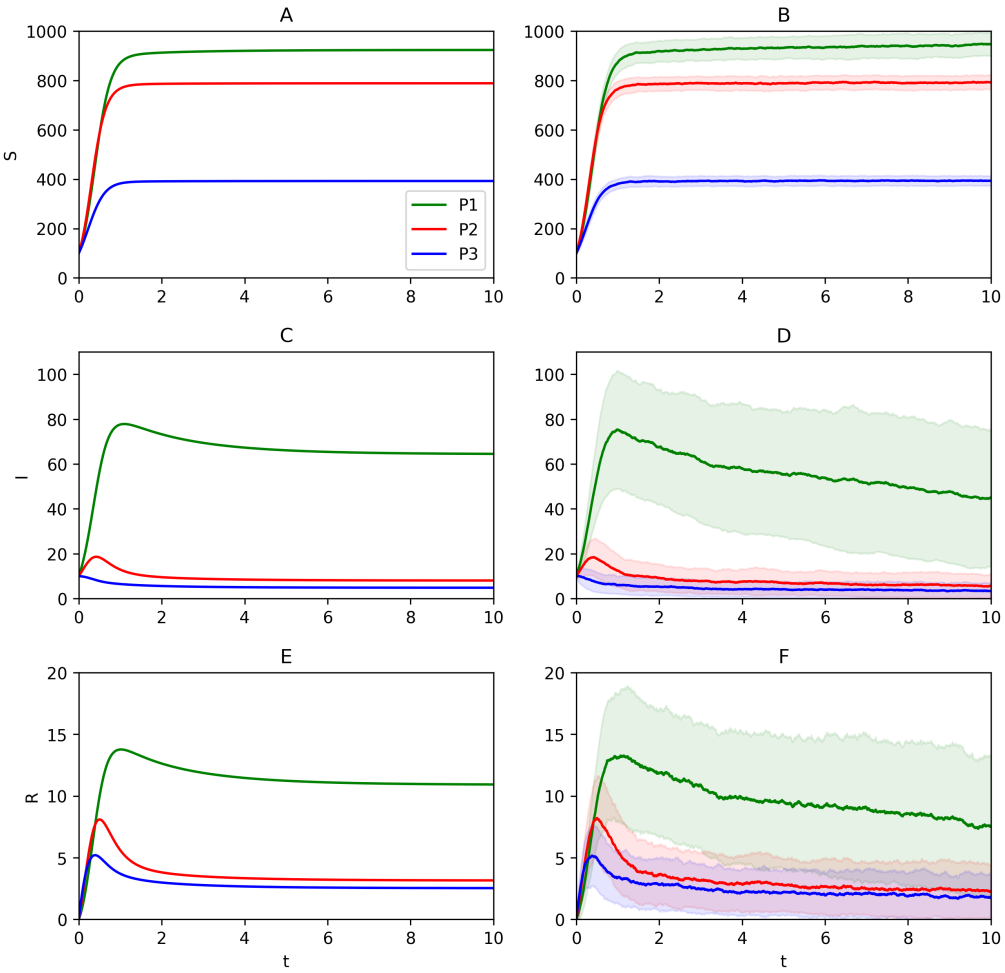


Figure 3.7: Side-by-side comparison of results of a multi-patch epidemic model created by differential equations from [157] (A, C and E) and *MetapopPy* (B, D and F). A and B show Susceptible population, C and D show Infectious population, E and F show Recovered population. Green plots are Patch 1, Red are Patch 2 and Blue are Patch 3. B, D and F show results from 500 *MetapopPy* simulations.

the equations of that model (a recreation of Figure 10 of [157], with plots grouped by compartment). The model is evenly coupled (all patches are connected are rate of movement is the same in both directions of each edge), and the differential parameters at each patch result in differing outcomes: the epidemic reaches a higher equilibrium in patch 1 than in patches 2 and 3.

In order to recreate this model in *MetapopPy*, a model-specific environment was created whereby each patch is defined with spatial attributes equivalent to the variable parameter rates (e.g. local birth rate, local death rate, local infection rate, etc.). Then, events were created which calculated their state variable using these environmental attributes. To demonstrate this, consider the infection event. In [157], this occurs at rate $S_j I_j \frac{\lambda_j}{N_j}$ at patch j. S_j, I_j and N_j are the Susceptible, Infectious and total populations at patch j, whilst λ_j is an infection rate specific to patch j. In *MetapopPy*, S_j, I_j and N_j are calculated dynamically during simulation (as they are constantly changing), whilst λ_j is a patch attribute. Thus assigning different values to the patch attributes creates different rates of infection across the environment[7].

The results from 500 simulations of the *MetapopPy* model are show in 3.7B, D and F. Again, we see that the results from a *MetapopPy* model are a suitable approximation of the results obtained from a deterministic equation-based model of the same scenario.

3.5 Conclusions

In this chapter, we have detailed the creation of a networked metapopulation framework, *MetapopPy*, designed to aid in the creation of metapopulation based models that include environmental heterogeneity. The framework provides the user with the underlying network data structure and event dynamics mechanisms, requiring the user only to define how the environment and events are described, without concern for the event and time processing algorithms. By extending *epyc*, we allow for experiments to be run across the model using a variety of parameter ranges, again removing the need for the user to define the experimental control mechanisms. We have verified our framework by recreating models from already-published sources and confirming that

[7] In order to make the rates valid, the reaction parameter for these events was set to 1. Another valid approach to this would be to set a reaction parameter equal to a standard infection rate and use patch attributes as a means of scaling the rates based on this value.

the output is similar. This framework then forms the basis of our TB whole-lung model, as described in the next chapter.

Chapter Four

Whole lung model of TB

In which we present TBMetapopPy, a novel whole-lung model of TB that incorporates environmental heterogeneity within the lung that impacts on the immunology and pathology of the disease dynamics, and present our results from the initial simulations.

4.1 Introduction

In previous chapters, we have explained how TB disease exhibits distinct localisations within the lung during different stages of disease, which is hypothesised to be caused by environmental heterogeneity within the lung. We have also shown that, to date, there are few computational models that explore TB disease occurring over the whole lung, and of those that do, none explore these differences in localisations or incorporate this concept of environmental heterogeneity within the lung which impacts on disease dynamics. In this chapter, we outline what we believe to be the first *in silico* model that simulates TB within an environmentally heterogeneous lung, in which the dynamics are influenced by differentials in environmental attributes such as ventilation, perfusion and oxygen tension, thus creating regions in the lung which are preferential or disadvantageous to bacterial proliferation.

We begin by providing an overview of anatomical and immunological concepts that are relevant to the creation of the model. We then outline the model, explaining how a differential environment is created by assigning values for environmental attributes based on each patch's position within the lung. We detail all of the events within the model, explaining their state variable calculations and the impact they have on the model and show how the environment influences the rates of some of these events. We then show the results of simulations using the model, demonstrating how the incorporation of environmental heterogeneity in a whole-lung model of TB can have significant impact on the outcome, with far greater bacterial proliferation occurring in a heterogeneous model due to the creation of regions at the apex whereby reduced immune activity and increased oxygen aid the cause of the bacteria.

4.2 Model description

We extended the *MetapopPy* framework presented in Chapter 3 to create *TBMetapopPy*, a model that simulates a TB infection over the whole lung, incorporating environmental heterogeneity and bacterial dissemination in order to understand the influences these factors have upon the progression of disease. As an extension of *MetapopPy*, *TBMetapopPy* follows the same class structure of Environment, Events and Dynamics. We discuss each of these in turn in the following subsections.

4.2.1 Environment

The lung is the main pulmonary organ of the human body. Air that is inhaled is directed into the lung through the trachea. This splits into two main bronchi, each supplying one lung. The right lung is larger in size than the left (as the left lung has to accommodate the heart within the chest cavity) and thus the main right bronchus is larger than the left. Each main bronchus divides into lobar bronchi, which each supply a lobe of the lung. The tissue of the lobes of the lung are separated by fissures. Due to size differences, the right lung contains 3 lobes whilst the left contains 2 [138]. Each of these lobes are also further subdivided

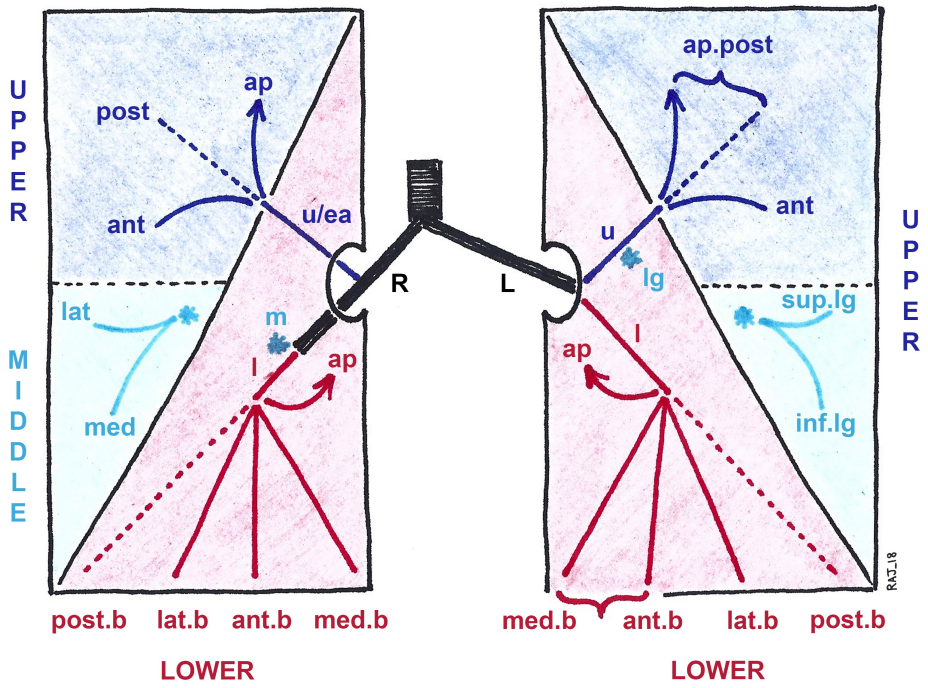


Figure 4.1: Diagram of the bronchopulmonary segments and lobes within the human lungs (taken from [24]). In the left lung, braces denote common fusions between segments.

into bronchopulmonary segment, with 10 bronchopulmonary segments in the right lung and 10 in the left, although in the left lung some fuse together [66]. It should be noted that lung anatomy is not constant, and differences in the number, size and shape of bronchopulmonary segment may occur between people. Each of these segments is supplied by its own segmental bronchus, branching off from the lobar bronchus (see Figure 4.1). A bronchopulmonary segment is the smallest functionally independent region of the lung [66].

Within each bronchopulmonary segment, the segmental bronchus splits (typically into no more than 2 subsequent airways) continuously, until the airways terminate at the alveolar sacs.

Our spatial domain for *TBMetapopPy* is a single lung and associated lymphatics and thus our model consists of two interlinked environments:

the alveolar tissue of the lung (LungPatch) and the lung-associated lymphatics (LymphPatch). Within the lung, the alveolar tissue is the region where most infection occurs, due to initial particle deposition there [37]. The average number of alveoli in an adult human lung is estimated to be in the range of 274–790 million [182], and these make up a combined surface area of approximately $70m^2$ [51]. We chose to model only a single lung for the initial model, and chose the right lung, as particle deposition (and thus TB primary disease) tends to occur more frequently in the right lung than the left, due to the size and angle of the main bronchi and lower airways [66].

The lymph nodes that serve the lung were chosen as they are necessary to trigger an adaptive immune response during infection [12], often show signs of disease during infection [16, 167] and are postulated as the means by which bacteria may disseminate to uninfected areas of the body [20, 119]. Thus, we used a 'TypedEnvironment' instance from the *MetapopPy* library, with patch types of 'LungPatch' and 'LymphPatch' to differentiate between the two organs.

4.2.1.1 Patch environmental attributes

In order to allow for a heterogeneous environment within the lung, each of the lung patches was assigned environmental attributes shown in Table 4.1. We chose attributes that have either been proven to impact on TB pathology or are assumed to do so (see section 2.9 for explanations on how each of these attributes impacts TB).

We made the assumption that these attributes, and only these attributes, have impact on TB progression. This included the following assumptions:

- The distribution of ventilation within the lung is linear, based on the patch's vertical position. Ventilation affects particle deposition, with lower patches more likely to receive an incoming particle than those higher up.

- The distribution of perfusion within the lung is also linear, and also

Label	Attribute	Description
V	Ventilation	The amount of inhaled air that is directed to the patch of alveolar tissue
Q	Perfusion	The amount of blood that is directed to the patch of alveolar tissue for oxygen exchange
O	Oxygen tension	The amount of oxygen remaining in the air within the lung after oxygen exchange has occurred
G	Drainage	The rate at which migratory cells within the lung translocate to the lymphatic system

Table 4.1: Environmental attributes included within the lung patches of *TBMetapopPy*

based on the patch's vertical position. Immune cell recruitment to the lung occurs through the blood used for oxygen exchange, and thus patches with a lower perfusion (towards the apices) will have lower levels of cell recruitment.

- The oxygen tension distribution is a direct result of the differentials for both ventilation and perfusion. Oxygen availability affects the replication capabilities of bacteria in the lung.

- Differentials in drainage are also linear in the lung, and also based on the patch's vertical position. Patches in a higher vertical position will have reduced rates of drainage than those towards the base.

Values for V, Q and G were defined at the beginning of simulation. The value for O was calculated automatically from these values, according to equation 4.1.

$$O = \frac{V}{Q} \tag{4.1}$$

The lymph patch was not assigned any environmental attributes as we are unaware of any environmental heterogeneity within the lymphatics that would impact TB pathophysiology.

Label	Compartment
B_{ER}	Bacterium extracellular - replicating
B_{ED}	Bacterium extracellular - dormant
B_{ID}	Bacterium intracellular - dendritic
B_{IM}	Bacterium intracellular - macrophage
D_I	Immature dendritic cell
D_M	Mature dendritic cell
M_R	Resting macrophage
M_I	Infected macrophage
M_A	Activated macrophage
T_N	Naïve T cell
T_A	Activated T cell

Table 4.2: Population compartments within *TBMetapopPy*

4.2.1.2 Patch population compartments

All patches were assigned populations divided into compartments, each of which represented the species and status of immune cells or bacteria, as described in Table 4.2.

We modelled 4 types of bacteria, based on their location and/or replication rate. Bacteria that are present on the tissue surface were 'extracellular'. *M. tuberculosis* has been shown to exhibit 'dormancy', associated with the accumulation of lipid bodies, whereby it reduces its replication rate but becomes more resistant to antibiotics [104, 142]. Thus we modelled this dormancy by allowing bacteria to switch between a 'replicating' (B_{ER}) and a 'dormant' (B_{ED}) state when extracellular, dependent on oxygen availability. *M. tuberculosis* has evolved to suppress the destructive mechanism of immune cells by preventing phagolysosome biogenesis, and thus is able to reside within the intracellular matrix of host cells [206, 225, 241]. We defined two types of intracellular bacteria - those within dendritic cells (B_{ID}) and those inside macrophages (B_{IM}), as both of these cells have been shown to become infected by *M. tuberculosis* [86]. We did not make the distinction between replicating and dormant when intracellular: we assumed that dendritic cells are too small to permit internal bacterial replication, and that the internal environment within a macrophage is hostile and thus forces the bacteria into a slower-replicating

state regardless of replication rate when ingested.

We modelled 3 types of immune cells: dendritic cells, macrophages and T cells. The primary role of dendritic cells in the model is antigen-presentation: the immature dendritic cells (D_I) resident in the lung encounter and ingest bacteria, causing the dendritic cell to convert to a mature state (D_M). These mature cells can then trigger an adaptive immune response by transferring to the lymphatics and activating T cells there.

Macrophages play a similar role in the model, as they can also encounter and ingest bacteria and transfer to the lymphatics, but we assumed that macrophages are less mobile than dendritic cells and have a greater internal capacity - they are more likely to remain in the lung and attempt to eliminate bacteria there. Thus, macrophages were given 'resting' (M_R) and 'infected' (M_I) states. We also included an 'activated' state (M_A), whereby the macrophage's bactericidal ability is improved during an adaptive immune response [75].

Naïve T cells (T_N) are present within the lymphatic system and may be activated by antigen-presenting cells (D_M and M_I). Multiple varieties of activated T cells exist within the lungs with varying different functions [78, 141]. For simplicity, we did not make a distinction between the different roles and include just one type of activated T cell (T_A) to serve as a representation for all real-world types.

4.2.1.3 Network topology

For our initial model, we chose to use a simple network in order to demonstrate that even a model with coarse environmental heterogeneity will exhibit profoundly different dynamics than one with a homogeneous domain. We chose to model the right lung (as this is where initial infection more often occurs) and base each of our patches on a bronchopulmonary segment: thus, the model contained 10 patches within the lung, and a single patch to model the adjacent lymph nodes. We used a single lymph

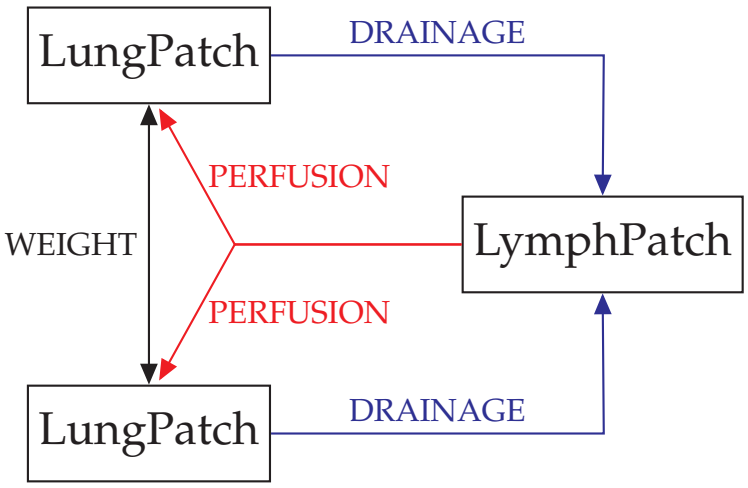


Figure 4.2: Translocation within the initial *TBMetapopPy* model.

patch to model the entire lymphatic system associated with the lungs.

The network is connected with edges, both between patches in the lung, and between lung patches and the lymph patch. Figure 4.2 shows the translocation opportunities within the network and the patch or edge attributes that affect translocation. Movement from a LungPatch to a LymphPatch is dependent on the G value of the LungPatch, whilst translocation from a LymphPatch to a LungPatch is dependent on the Q value of the LungPatch[1]. Doing so allows for a differential T cell response: T cells in the LymphPatch can transfer into the lung, with a greater rate translocating to patches with a higher Q value.

We fully connected all patches within the lung, using the fact that the internal structure of the lung is a tree to assume that all areas are connected and any bacteria in one area will be able to transfer bronchogenically (see section 2.8) to any other area. We assigned a weight (W) value to each edge. We assumed that patches within each lobe are closer on the bronchial tree than those in other patches, and thus the W value is higher within-lobe than between-lobe.

The overall topology of the network can be seen in whole in Figure 4.3.

[1]In order to achieve this, the Q value of the LungPatch was also assigned to the edge connecting the LungPatch and LymphPatch, thus allowing events at the LymphPatch to view the perfusion value and calculate state variables using it

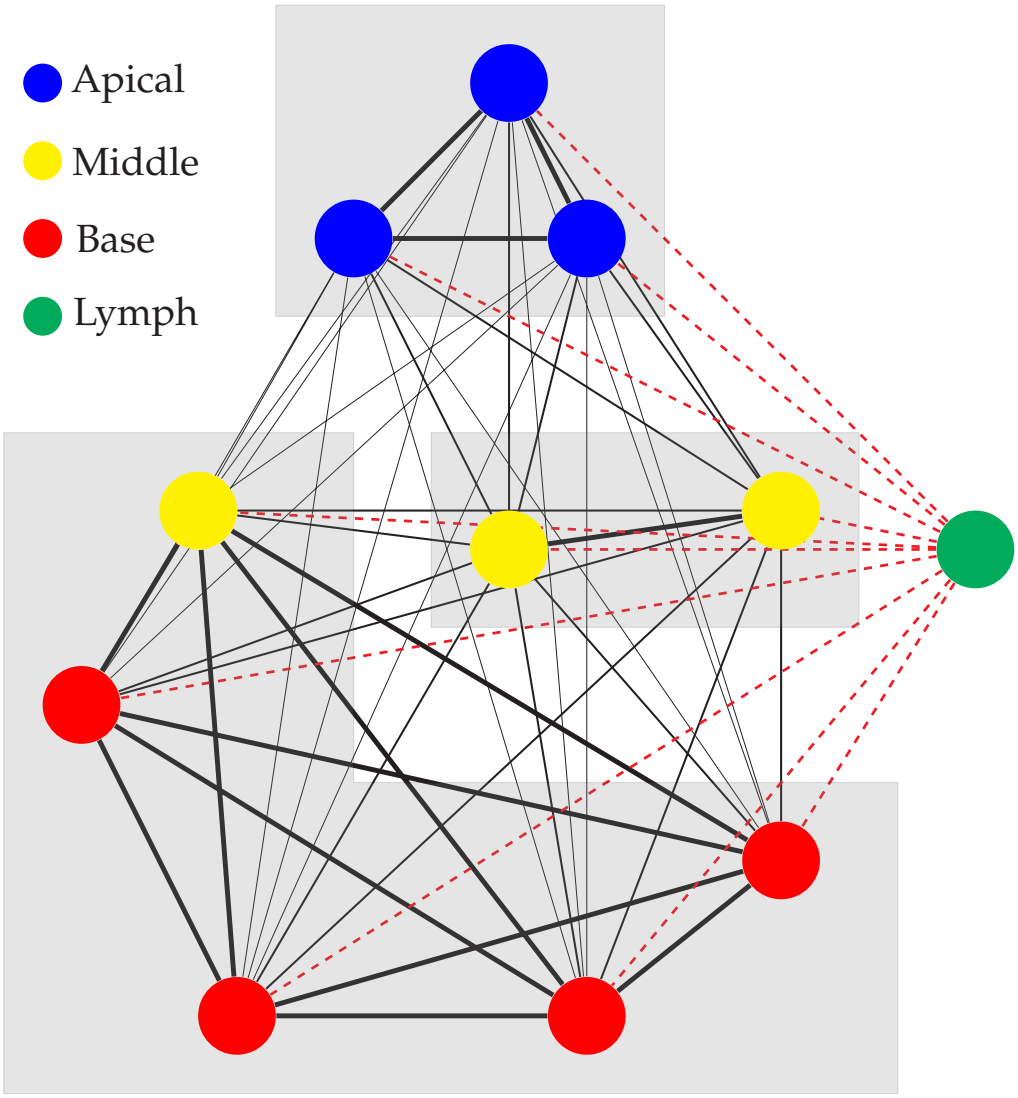


Figure 4.3: Network topology for initial *TBMetapopPy* model.

The 10 patches of the lung (based on bronchopulmonary segments) are placed into lobes based anatomically, with the upper lobe containing 3 patches, the middle lobe containing 2 patches. and the lower lobe containing 5 patches. In order to define values for environmental heterogeneity, we separate the lung into 3 regions: apical, middle and base (blue, yellow and red, respectively, in Figure 4.3). [233], defined three distinct zones within the lung based on magnitudes of pulmonary arterial, venous and alveolar blood pressures, and the zones in our model can be seen as analogous to these zones. All patches within the same zone were assumed to have the same values for ventilation and perfusion (and thus oxygen).

4.2.2 Events

Here, we describe the events that occur within the initial *TBMetapopPy* model. We describe each event in terms of its action (i.e. the changes made by its perform function) and its rate (i.e. its reaction parameter multiplied by its state variable). Figure 4.4 shows a schematic of how events move members between compartments and a list of all parameters can be found in Tables 4.3, 4.4, 4.5, and 4.6.

4.2.2.1 Innate immune response

M_R cells are initially present in both the lung and lymphatic system. These cells die naturally at a rate of μ_{MR} and are replaced through recruitment, which occurs at rate $Q\alpha_{ML}$ in the lung and α_{MY} in the lymph. D_I cells are also present in the lung at simulation start; these die at rate μ_{DI} and are replaced through recruitment at rate $Q\alpha_D$. In the lymphatics, naïve T cells are present at the beginning of simulation; these die at rate μ_{TN} and are replaced through recruited cells at rate α_T.

The bacteria in the lung are initially extracellular and are able to replicate freely, with B_{ER} replicating at rate $\lambda_R B_{ER}$ and B_{ED} replicating at rate $\lambda_D B_{ED}$. In order to model the effects of hypoxia on bacterial replication, we include an event to switch bacteria from B_{ER} to B_{ED} with increased likelihood in an oxygen-poor environment, with rate $\xi B_{ER}\left(\frac{O^{-\sigma_\xi}}{O^{-\sigma_\xi}+\theta_\xi{}^{-\sigma_\xi}}\right)$,

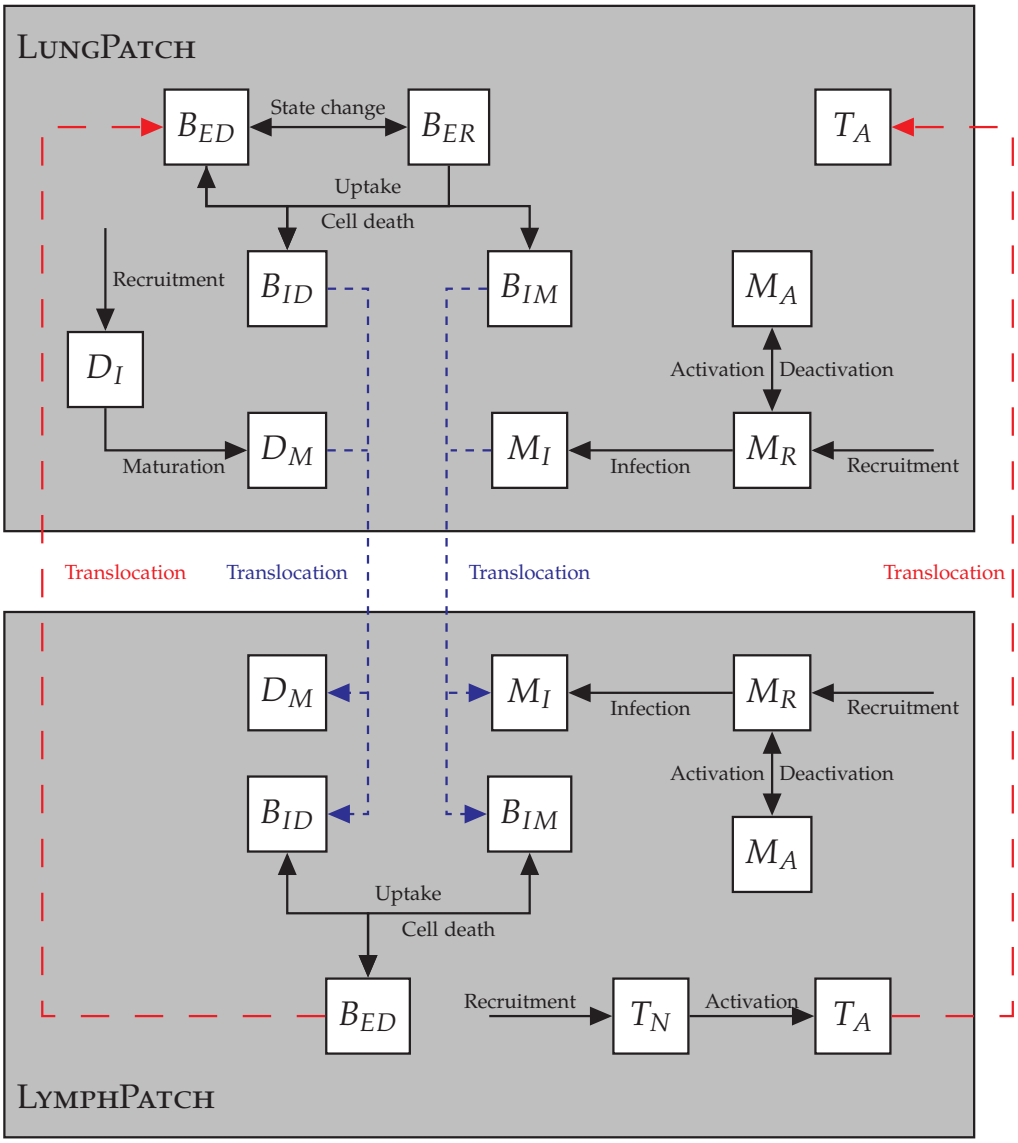


Figure 4.4: Schematic showing the movement of members between populations (and patches) within the preliminary *TBMetapopPy* model. Translocation of extracellular bacteria between LungPatch instances is also included in the model (not shown here).

Parameter	Description	Value	Ref
ϕ	Bacterial carrying capacity of a macrophage	50	[153]
σ_λ	Intracellular bacteria replication sigmoid	2	[153]
λ_R	Replication rate of B_{ER}	0.25	[240]
λ_D	Replication rate of B_{ED}	0.005	[240]
λ_I	Replication rate of B_{IM}	0.005	Estimated
κ_{BR}	Destruction rate of B_{ER} and B_{ED} by M_R	1.25E-8	[153]
κ_{BA}	Destruction rate of B_{ER} and B_{ED} by M_A	1.25E-4	Estimated
ξ	Rate of conversion between B_{ED} and B_{ER}	1	Estimated
θ_ξ	Half-sat value of O for conversion between B_{ER} and B_{ED}	1	Estimated
σ_ξ	Sigmoid value for conversion between B_{ER} and B_{ED}	1	Estimated
τ_{BL}	Rate of translocation of B_{ER} and B_{ED} along LungEdge (within lung)	0.001	Estimated
τ_{BY}	Rate of translocation of B_{ED} along BloodEdge (lymph to lung)	0.05	Estimated

Table 4.3: Bacterial event parameters used for experiments in the preliminary *TBMetapopPy* model. All value units are per day.

whereby $\sigma_\xi > 0$. We also include the reverse event, with B_{ED} converting to B_{ER} when presented with an oxygenated area, with rate equal to $\xi B_{ED}\left(\frac{O^{\sigma_\xi}}{O^{\sigma_\xi}+\theta_\xi{}^{\sigma_\xi}}\right)$.

The resident phagocytes (M_R and D_I) within the lung may encounter the extracellular bacteria and ingest them. For dendritic cells, we assumed that a dendritic cell is incapable of destroying a bacteria and can only uptake them. This occurs at rate $\gamma_D D_I \frac{B_E}{B_E+\theta_{\gamma D}}$, where $B_E = B_{ER} + B_{ED}$. When performed, an extracellular bacteria is probabilistically chosen to be ingested based on current levels. For macrophages, we modelled *M. tuberculosis*'s ability to avoid destruction through two events: one where the

Parameter	Description	Value	Ref
μ_{DI}	Rate of death of D_I	0.01	[153]
μ_{DM}	Rate of death of D_M	0.02	[153]
γ_D	Rate of uptake of B_{ER} and B_{ED} by D_I, causing maturation	1E-7	[153]
α_D	Standard recruitment rate of D_I into lung	500	[153]
β_D	Enhanced recruitment rate of D_I into lung caused by B_{ED} and B_{ER}	0.02	[153]
$\theta_{\beta D}$	Half-sat for enhanced recruitment rate of D_I into lung caused by B_{ED} and B_{ER}	6.5E5	[153]
τ_D	Rate of translocation of D_M along LymphEdge (lung to lymph)	0.5	[153]

Table 4.4: Dendritic cell event parameters used for experiments in the preliminary *TBMetapopPy* model. All value units are per day.

bacteria is destroyed upon contact, occurring at rate $\kappa_{BR} M_R B_E$, and one where the macrophage becomes infected, occurring at rate $\pi M_R \frac{B_E}{B_E + \theta_\pi}$. For both these events, the bacterium is chosen probabilistically based on current extracellular levels of B_{ER} and B_{ED}. We assumed that the different replication phenotypes (B_{ER} or B_{ED}) have the same chances of survival when encountered by macrophages.

If a bacteria survives the ingestion process, it converts to an intracellular state (B_{IM} or B_{ID} for phagocytosis by M_R or D_I respectively) and the cell converts to an infected state (M_R convert to M_I, D_I convert to D_M). M_I and D_M both die naturally, at rates $\mu_{MI} M_I$ and $\mu_{DM} D_M$ respectively, and this may return some of the internal bacteria back into extracellular compartment B_{ED} (we assume that the internal conditions within a macrophage are stressful to bacteria and thus they have been forced into a dormant state, which they remain in once released). For D_M death, we assumed the single internal B_{ID} is always released, whilst for M_I death, we assumed a percentage of the bacteria inside the macro-

Parameter	Description	Value	Ref
ϵ_{MT}	Rate of activation/deactivation of M_R by T_A	0.3	[153]
$\theta_{\epsilon MT}$	Half-sat value for activation of M_R by T_A	5E2	Estimated
π	Rate of macrophage infection by B_{ER} and B_{ED}	0.4	[153]
θ_{π}	Half-sat value for macrophage infection	1E6	[153]
μ_{MR}	Death rate of M_R	0.01	[153]
μ_{MA}	Death rate of M_A	0.01	[153]
μ_{MI}	Standard death rate of M_I	0.01	[153]
ζ	Rate of bursting of M_I due to B_{IM}	0.25	[153]
α_{ML}	Standard rate of recruitment of M_R into lung	400	[240]
α_{MY}	Standard rate of recruitment of M_R into lymph	53.465	[154]
β_{MLI}	Enhanced rate of recruitment of M_R by M_I	0.0056	[153]
β_{MLA}	Enhanced rate of recruitment of M_R by M_A	0.04	[153]
τ_M	Rate of translocation of M_I along LymphEdge (lung to lymph)	0.25	Estimated

Table 4.5: Macrophage event parameters used for experiments in the preliminary *TBMetapopPy* model. All value units are per day.

phage ($\psi_\mu \frac{B_{IM}}{M_I}$) are destroyed with the remainder ($(1-\psi_\mu)\frac{B_{IM}}{M_I}$) released.[2]

Once in the internal environment of the immune cell, bacteria can replicate [197]. We allow replication of B_{IM} to occur, constrained by the internal capacity of the cell. B_{IM} may replicate at rate $\lambda_I B_{IM}\left(1-\frac{(B_{IM})^{\sigma_\lambda}}{(B_{IM})^{\sigma_\lambda}+(\phi M_I)^{\sigma_\lambda}}\right)$. Thus the rate of replication decreases as the levels of bacteria approach the total carrying capacity of all infected macrophages (ϕM_I). As macrophages start to fill with bacteria, they may 'burst', rupturing their outer wall and releasing bacteria [62]. The rate of occurrence for this

[2]These values in *TBMetapopPy* (and all other compartment values where fractions are calculated) are rounded to whole numbers in order to comply with *MetapopPy* rules that compartment values must be integers.

Parameter	Description	Value	Ref
κ_M	Rate of destruction of M_I by T_A	1.3	[153]
$\theta_{\kappa M}$	Half-sat value for destruction of M_I by T_A	0.5	[153]
ϵ_T	Rate of T-cell activation by antigen presentation by M_I and D_M	1E-5	[153]
μ_{TN}	Rate of death of T_N	0.102	[153]
μ_{TA}	Rate of death of T_A	0.3333	[153]
α_T	Standard rate of recruitment of T_N into lymph	1000	[153]
β_T	Enhanced rate of recruitment of T_N into lymph due to D_M and M_I	0.1	Estimated
τ_T	Rate of translocation of T_A along BloodEdge (lymph to lung)	0.6	[153]

Table 4.6: T-cell event parameters used for experiments in the preliminary *TBMetapopPy* model. All value units are per day.

in *TBMetapopPy* is inverse to replication, i.e. it increases as the bacterial level approaches capacity, with rate $\beta M_I \left(\frac{(B_{IM})^{\sigma_\lambda}}{(B_{IM})^{\sigma_\lambda} + (\phi M_I)^{\sigma_\lambda}} \right)$. When this occurs, the internal bacterial contents of the macrophage are return to the extracellular compartment, which is calculated as $\frac{B_{IM}}{M_I}$ members switching from B_{IM} to B_{ED}. We assume D_I cells are too small to permit bacterial replication within themselves and thus their internal bacteria, B_{ID}, do not replicate.

Extracellular bacteria may translocate: bacteria present in a LungPatch may translocate along the air (via LungEdge instances) at rate $\tau_{BL} B_{ER}$ and $\tau_{BL} B_{ED}$ for B_{ER} and B_{ED} respectively. The choice of edge depends on the weighting value of the edge, with greater probability along edges with greater W values. Extracellular bacteria in a LymphPatch may translocate along a BloodEdge back into the lung at rate $\tau_{BY} B_{ED}$, with the edge to move along chosen probabilistically based on the Q value.

4.2.2.2 Adaptive response

In response to infection, the body increases the supply of immune cells to the lung to assist within containment [93]. This is modelled in *TBMetapopPy* by separate events for each resident immune cell and trigger cell type at each location. We assume that signalling cytokines are expressed by M_I and M_A cells, with enhanced recruited of macrophages due to each cell occurring at rates $\beta_{MLI}M_I$ and $\beta_{MLA}M_A$ respectively in both the lung and lymph patches. In the lung, dendritic cell recruitment is enhanced by bacterial presence at rate $\beta_D \frac{B_E}{B_E+\theta_{\beta D}}$. In the lymphatics, enhanced T cell recruitment is dependent on APCs and is increased at rate $\beta_T(D_M + M_I)$.

Both dendritic cells [6, 164] and macrophages [130] have been shown to transfer to the lymphatics during infection in order to present antigens and trigger an adaptive immune response. We treat both D_I and M_I as antigen-presenting cells and we model their transfer to the lymphatics as two separate events: with D_I translocating at rate $\tau_D G D_I$ and M_I translocating at rate $\tau_M G M_I$, thereby creating different rates of clearance at different regions, based on G. This transfer, whilst necessary for establishing an adaptive immune response, also carries the risk of spreading infection, as any internalised bacteria are also transferred to the lymphatics. The transfer of bacteria to the lymph nodes may be a necessity for establishing an adaptive immune response [242]. Thus in our model, transfer of M_I also causes the transfer of B_{IM} members (calculated as the average number of B_{IM} per M_I, $\frac{B_{IM}}{M_I}$), and transfer of a D_M member also transfers a single B_{ID}.

Antigen-presenting cells in the lymphatics can cause activation of T cells from a naïve to an activated state. This occurs at rate $\epsilon_T T_N(M_I + D_M)$, and results in T_N changing to T_A. Once activated, these T cells migrate to the sources of infection to enhance the immune response and contain the bacteria. The rate of translocation is $\tau_T T_A$. When choosing a LungPatch to translocate to, the Q values of each patch serve as a means of probabilistically choosing a patch: patches with a greater perfusion value are

more likely to be chosen and thus a greater T cell response is initiated at patches with greater perfusion.

Activated T cells perform two primary functions in the model. The first is to cause activation of macrophages, converting M_R into M_A, occurring at rate $\epsilon_{MT} M_R \frac{T_A}{T_A + \theta_{\epsilon MT}}$. The second function of T_A is to causes apoptosis - the controlled destruction of infected cells. This is modelled by allowing T_A members to destroy M_I members at rate $\kappa_M M_I \frac{T_A/M_I}{(T_A/M_I) + \theta_{\kappa M}}$. We assumed that the destuction of the macrophage also destroyed the internal bacteria, calculated as $\frac{B_{IM}}{M_I}$, which are removed from the B_{IM} compartment. T_A members naturally die at rate $\mu_{TA} T_A$

Activated macrophages can destroy bacteria in the same manner as M_R, with rate $\kappa_{BA} M_A B_E$, but in our model M_A cannot become infected and thus the bacteria ingested are always destroyed, simulating the increased bactericidal effects of macrophage activation. M_A cells die naturally at rate $\mu_{MA} M_A$.

4.2.3 Dynamics

A TBModel class (an extension of the Dynamics class of *MetapopPy*) combines the pulmonary environment and TB events.

4.2.3.1 Initial conditions

The lung topology is initialised and seeded with values for resident immune cells (M_R and D_I in the LungPatches, M_R and T_N in the LymphPatch) based on the recruitment and death values to calculate average equilibrium levels of these cells. A population of bacteria are seeded, either in a defined location or probabilistically based on Q values. The simulation proceeds as per the *MetapopPy* framework until a simulated time limit is reached.

4.3 Results

4.3.1 Set-up

In order to demonstrate the effects of environmental heterogeneity upon TB infection within the lungs, we performed experiments against two distinct topologies of the *TBMetapopPy* model. In the first, topology A, the lung is modelled as a whole homogeneous unit, consisting of just one LungPatch coupled to a LymphPatch which models the lymphatics. A LymphEdge and a BloodEdge allow movement of cells between the two. The second, topology B, is the setup as described in section 4.2.1.3. wed defined attribute values for each zone as per Table 4.7.

	V	Q	O
Apical	0.6	0.5	1.2
Middle	0.8	0.8	1.0
Basal	1.2	1.3	0.923

Table 4.7: Values for attributes of West Zones used in simulations of multi-patch topology model. V = Ventilation, Q = Perfusion, O = Oxygen tension.

As the model is stochastic in nature, all experiments are run over 30 repetitions, with the mean counts of members (and standard deviations) of each compartment over these repetitions calculated at each time-step. For experiments involving topology B, 5 B_{ER} members were placed into one of the basal regions at the start of simulation. In topology A models, they are placed in the single lung patch.

We are interested in overall bacterial growth, and our marker for this is the combined total of the B_{ER}, B_{ED}, B_{IM} and B_{ID} compartments in each patch of the lung.

4.3.2 Topology A: Homogeneous systems require significant reduction in immune activity in order to establish infection

Figure 4.5a) shows the overall bacterial growth over time for the model using topology A and parameters as per Tables 4.3, 4.4, 4.5 and 4.6. For this

model, we set the V and Q values of the patch to 9.0 and 9.1 respectively, chosen in order to match the sum of the total ventilation and perfusion over an entire lung as seen in topology B. Bacterial growth is severely limited, as the bacteria population is overwhelmed by the numbers of immune cells, and the entire population of bacteria is eradicated within 750 days.

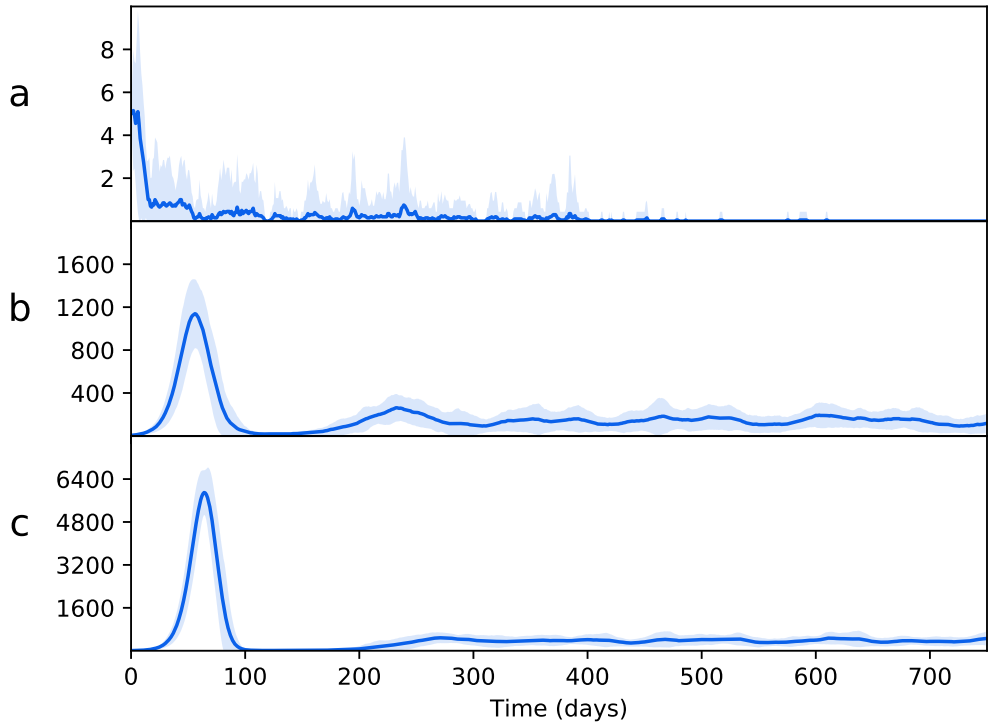


Figure 4.5: The overall bacterial growth (B_{ER}, B_{ED}, B_{IM} and B_{ID}) within a single patch topology (Topology A). Line is mean, shading is standard deviation. a) shows results using event parameters as described in Tables 4.3, 4.4, 4.5 and 4.6. Infection is contained quickly and virtually eradicated. b) shows the effects of a minimal reduction in cell recruitment (R_D = 50, R_{M_lung} = 40) and c) shows the effects of a major reduction in cell recruitment (R_D = 25, R_{M_lung} = 20). As recruitment is reduced, the spike of initial infection (at around 50 days) increases massively. In both cases, the infection is contained and remains at a reasonably steady state thereafter (latency).

We then investigated the conditions necessary to establish a bacterial presence. In order to do this, we reduced the recruitment rates of dendritic cells and macrophages (α_D and α_{ML} respectively) by arbitrary factors

of 10 (shown in Figure 4.5B) and then 20 (shown in Figure 4.5C). This severe weakening of the immune system recruitment throughout the course of infection is enough to allow TB to establish a presence within the lungs.

This was explored further by instead reducing the Q value of the lung patch in topology A. This has the effect of also weakening the adaptive immune response, which requires perfusion in order to increase macrophage and dendritic cell numbers and allow T cells to access the lung. Figures 4.6A and B show the effects of reducing the Q value of the single patch to 4.5 and 1.3 respectively. Reducing perfusion allows the establishment of a bacterial presence, with a higher bacterial load being established with perfusion reduced to very low levels. In Figure 4.6c), perfusion of the whole lung is reduced to levels used in the apical regions of topology B (i.e. 0.5). Here, a bacterial presence is firmly established. This suggests that lack of perfusion may be the main cause for bacterial growth as low-perfused patches experience high bacterial growth in both topologies.

4.3.3 Topology B: Spatial heterogeneity increases overall bacterial load

In order to explore the effects of environmental heterogeneity (both in the separation of the overall lung population into discrete patches and the variation of resources within those patches) and translocation on these disease dynamics, we ran simulations over the spatially separated network of topology B. Figure 4.7a) shows the results of simulation runs where no attribute heterogeneity exists (all patch values are set equal). From this, we see that the separation into distinct spatial regions results in much greater initial bacteria growth, due to the limitations on the number of immune cells that can encounter the bacteria during the early stages on infection. An adaptive immune response begins to quell the infection, but in doing so it results in transfer of the bacteria to other areas of the lung. An oscillating process of bacterial growth occurs, with bacteria numbers increasing and decreasing over time, but never reaching the levels of the initial growth explosion. This can be seen as a latent

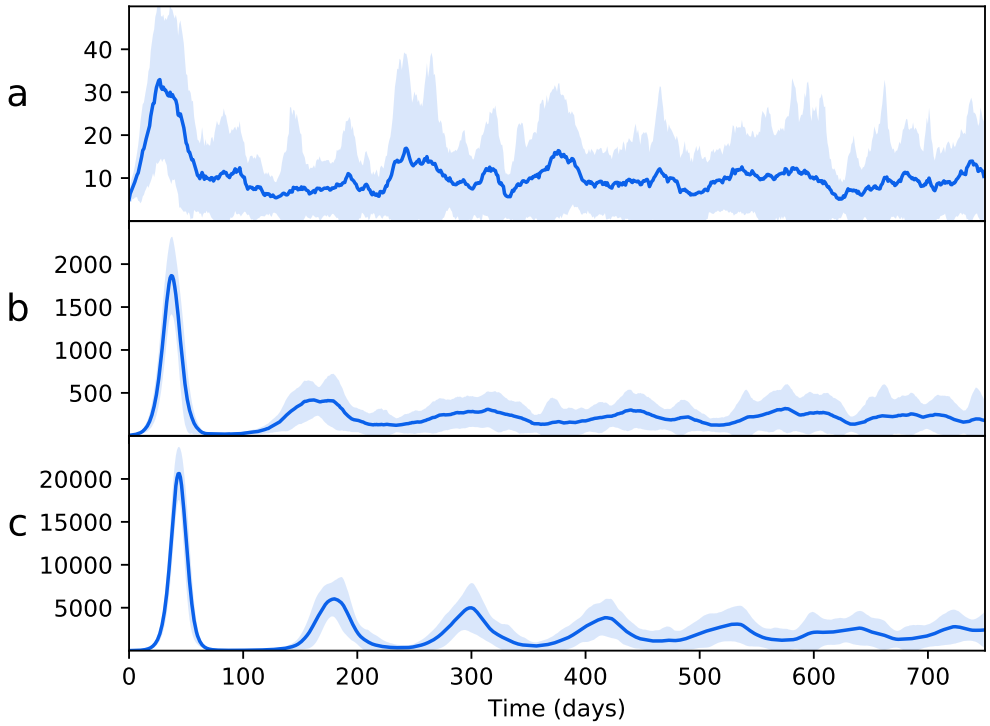


Figure 4.6: The overall bacterial growth (B_{ER}, B_{ED}, B_{IM} and B_{ID}) within a single patch topology (Topology A) when perfusion is reduced to a) 4.5, b) 1.3, c) 0.5. Line is mean, shading is standard deviation. Reducing perfusion allows a bacterial presence to be established. In a) no major initial infection takes place and the infection level remains fairly constant throughout. In b) a pattern similar to the reduced recruitment rate in Figure 4.5 is seen. In c), the reduced perfusion results in a saw-tooth pattern, whereby latency now involves an oscillating, rather than constant, level of bacteria.

form of TB: the bacteria levels are contained but not eradicated. As all patches contain the same values for their attributes, the bacteria levels at each patch remain approximately the same. The bacterial levels here far outweigh those seen in the single-patch model, except in the case where perfusion is drastically reduced.

We then introduced spatial heterogeneity to the different zones, with values as per Table 4.7, the results of which are shown in Figure 4.7b). As per the previous experiment, bacterial growth occurs in the initial patch until the introduction of an adaptive immune response, which brings the bacterial levels under control but at the consequence of spreading bacteria

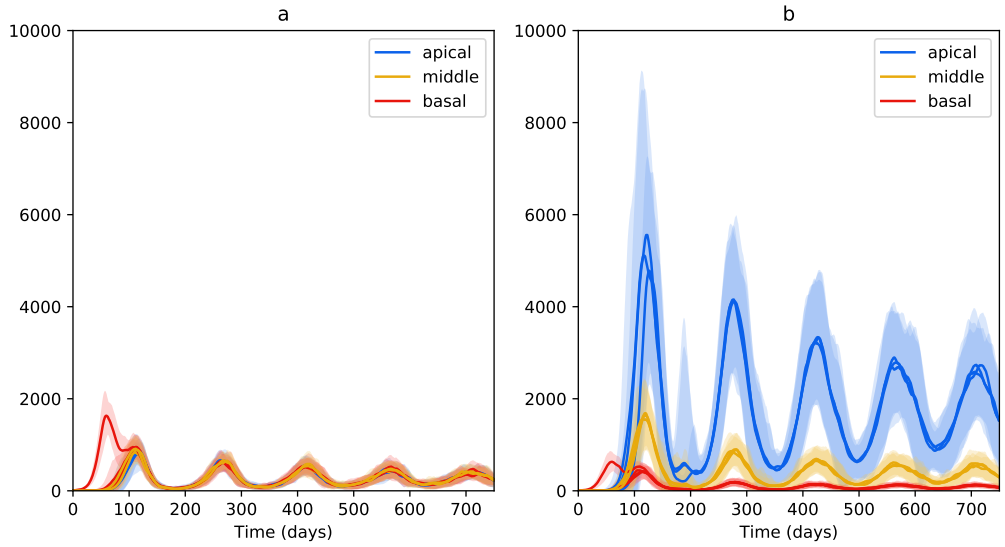


Figure 4.7: The overall bacterial growth (B_{ER}, B_{ED}, B_{IM} and B_{ID}) upon a multiple patch topology. Patches are grouped by spatial location (base, middle and apex). Line is mean, shading is standard deviation. All simulations use event parameters as per Tables 4.3, 4.4, 4.5 and 4.6. a) shows the results of a network that includes no spatial heterogeneity - initial infection spikes in the initial, basal patch and spreads to other patches, where infection is established at consistent levels throughout. b) shows the effects of introducing heterogeneity (with values as per Table 4.7). Infection spreads to other patches, with the bacterial load dependent on the spatial location.

to other regions. However, in this scenario the bacteria have access to the oxygen-rich, immune-poor regions at the lung apices, and once there the population explodes dramatically, reaching levels far higher than it did at the basal regions during initial infection. Again, the adaptive immune response is able to bring the bacterial levels down and reach an oscillating equilibrium. But in this scenario, we show that the levels of bacteria at latency are different for each region, with far higher levels being established at the apical regions than at the basal and middle regions. Figure 4.8 shows the levels of replicating and dormant bacteria in each region of the multi-patch network. We see that there are similar levels of replicating and dormant bacteria, with a slight majority to replicating bacteria in all patches.

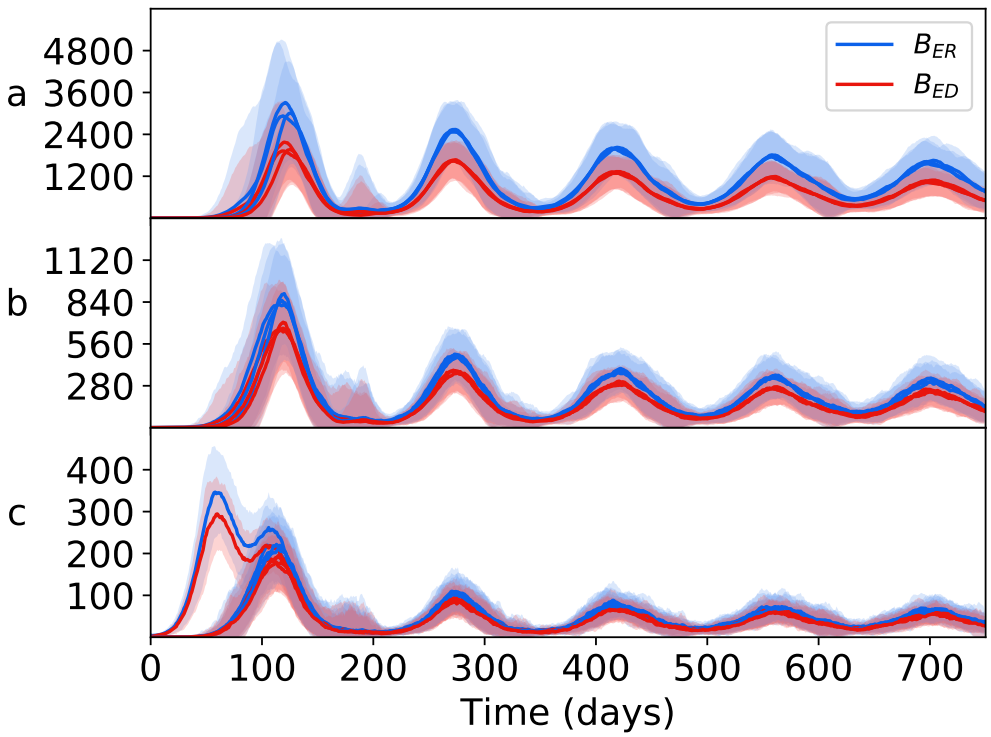


Figure 4.8: The bacterial levels of B_{ER} and B_{ED} in a) apical patches, b) middle patches and c) basal patches. Line is mean, shading is standard deviation. All three regions show similar levels between replicating (blue) and dormant (red) bacteria, with a slight majority to replicating bacteria in all.

4.3.4 Bacterial translocation

To investigate the effects that translocation within the lung had on bacterial growth, we ran experiments using different weight (W) values on the LungEdges connecting the patches within topology B. Figure 4.7b) shows the effects of no translocation within the lung (all W values set to 0), whilst Figure 4.7c) shows the effects with translocation allowed between patches which reside in the same lobe (W attribute of edges between patches in the same lobe is set to 1.0, and 0.0 for all other edges). Figure 4.7d) shows the effects of allowing translocation between all areas of the lung. W values are set as per Table 4.8.

In Figure 4.9, we tested the effects of removing a) long-scale inter-lung transmission b) all inter-lung transmission and c) bloodstream

	superior	middle	inferior
superior	1.0	0.5	0.25
middle	-	1.0	0.5
inferior	-	-	1.0

Table 4.8: WEIGHT values for edges between patches in different lobes

transmission from the lymphatics to the lung. In all cases, the overall pattern of bacterial growth in the lungs remains fairly robust, with minor changes in the amplitude of the resulting oscillations of bacteria levels. These findings suggest that the means of translocation across the pulmonary system are not as important as the heterogeneity therein. So long as bacteria can reach the favourable apical locations in some manner, they will be able to proliferate and establish a greater infection there.

4.4 Analysis

4.4.1 The need for environmental heterogeneity when modelling TB

It is believed that the environmental heterogeneities present in the lung are responsible for the apical localisation seen during the post-primary stage of disease [67], and the results from our preliminary model would seem to validate this hypothesis. By including differentials within the lung for attributes such as ventilation, perfusion and oxygen tension that mirror those evidenced in human anatomy, our model created regions of the lung at the apex that were preferential for bacterial growth. At these regions, the bacterial numbers far exceeded the numbers that occurred in a similar model without environmental heterogeneity. This could explain the localisation at the apex during post-primary disease: when latency is established, the lesions at the apical regions contain greater bacilliary loads, and thus under a weakening of the immune system, it is reasonable to assume that these lesions would be the first to degrade and experience bacterial proliferation on a rapid scale. This hypothesis is tested in later chapters of this thesis.

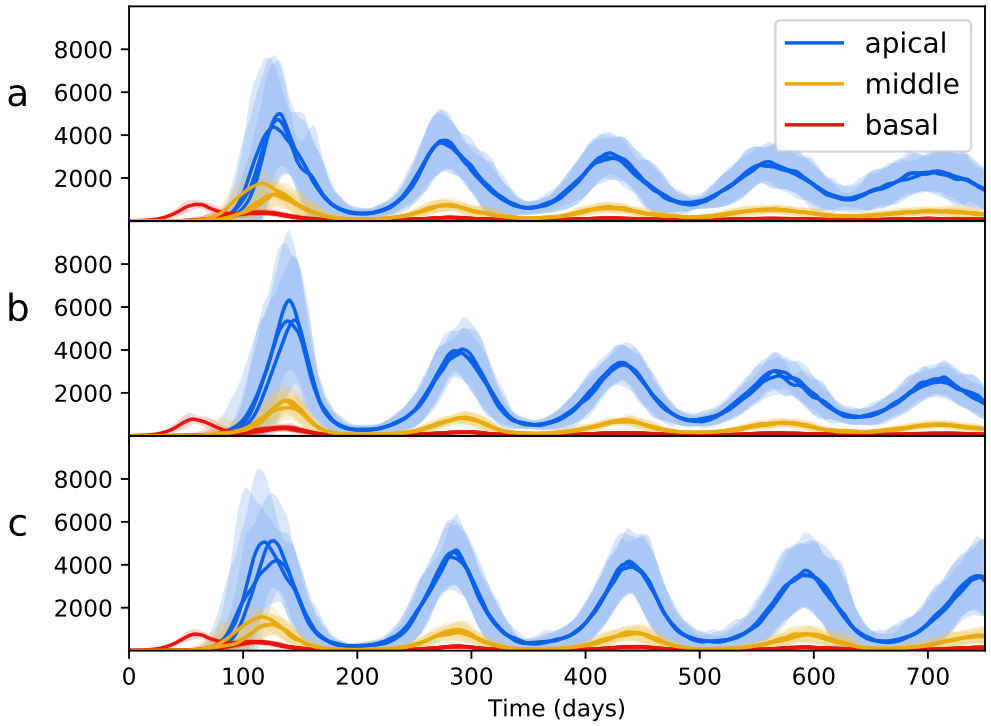


Figure 4.9: The overall bacterial growth (B_{ER}, B_{ED}, B_{IM} and B_{ID}) upon a multiple patch topology with different routes of bacterial dissemination. Patches are grouped by spatial location (base, middle and apex). Line is mean, shading is standard deviation. All simulations use event parameters as per Tables 4.3, 4.4, 4.5 and 4.6. a) shows the results of a network where translocation within the lung can only occur within the same lobe (i.e. small-distance translocation). b) shows the effects of completing removing translocation within the lung. c) shows the effects of removing translocation of bacteria via the lymphatics and bloodstream back to the lung. In all cases, removing options of translocation results in a similar pattern of growth to when all methods are viable (as per Figure 4.7).

The differences in routes of bacterial translocation across the pulmonary system appear to have little effect on the disease outcome; so long as bacteria have some means of reaching the preferential apical regions (i.e. through dissemination in the blood after establishing lymphatic disease or via direct transfer within the air of the lung), this will be enough to cause an enhanced population growth of bacteria in the region due to its favourable environment.

Figure 4.8 indicates that the bacterial population in all regions is composed of both replicating and dormant bacteria in almost equal measure. Whilst it is believed that a latent form of infection involves the host maintaining a population of dormant, non-replicating bacteria [181], there is evidence that some bacterial replication occurs during latency [116, 128], which may explain the presence of a population of B_{ER} bacteria after initial infection and the oscillating bacteria levels in our model.

With the results presented in this chapter, we have demonstrated the necessity of *in silico* modelling of TB infections to take a broader view of the disease as it occurs within the lung than is done currently: environmental factors play an important role in determining the scale of infection before latency is established. Developing new treatment regimens requires a better understanding of exactly which environments bacteria persist in within the lungs - computational models such as the one introduced here can help shed light on these niches.

4.4.2 Limitations

This model is a highly abstracted representation of the actual conditions found during a TB infection, but it is nevertheless based on gross-scale real world pathology. These abstractions are necessitated by the current lack of understanding of how the disease pathology is influenced by its surroundings. We have shown that the introduction of a small amount of heterogeneity, by dividing the lung into a network consisting of 10 nodes, can have significant impact on the results of simulations in contrast with models that treat the lung as a homogeneous environment.

The purpose of this model is not, presently, to be predictive; it serves to provide insight into the complex factors that drive TB infections to their apical localisation. This insight can provide the basis for models which can build on this work to further study the impact of lung environment on TB. We have identified some parameters whose spatial heterogeneity is significant to apical localisation (such as reduced perfusion at the apices), but others may exist. Where possible, we have used parameters taken

directly from the scientific literature.

The model presented in this chapter uses the environmental oxygen level of the patch as the only influence on the switch between replication and dormancy - in reality, a wide range of factors are believed to initiate dormancy, including oxygen deficiency, nutrient deficiency, hyper-acidity and antibiotics [142]. These stresses may drive bacteria to dormancy all over the lung, with the oxygen availability at the apices then providing the ideal environment for future reactivation.

The pathophysiology of TB is a complex process involving many and various interactions between the bacteria and a large number and variety of immune cells. In this model, we have simulated just three types of immune cells with a small number of states. In reality, there are many more immune cells involved in the progression of the disease, including neutrophils [71, 168], which are believed to play a significant role in the pathogenesis of TB as they are one of the first cells to encounter bacteria [145]; and B cells [1, 146]. For this iteration of the model, the cells we have included are sufficient to show that heterogeneity in the lungs can impact immune response and thus also disease pathology, but future iterations that intend to fully explore the immune response should incorporate these types of cells.

This preliminary *TBMetapopPy* model is an approximation of two stages of the TB life-cycle: primary infection and latency. In order to better understand how environmental factors contribute to disease, the model would need to incorporate factors appropriate to post-primary disease, as this is the stage that results in disease transmission and patient mortality. This is explored in the following chapter.

Chapter Five

Full life-cycle model of TB

In which we present the second iteration of TBMetapopPy, where a space-filling tree algorithm is used to create a more granular network, and present our results from simulations of this model, showing how position in the lung affects lesion development.

5.1 Introduction

As noted in the previous chapter, the preliminary *TBMetapopPy* model effectively models the primary and latency stages of TB. However, almost all transmission and the majority of mortality of disease (80%) occurs in the post-primary stage [125], typically localised to the apices of the lung and displaying cavitation - which greatly increases the chance of disease transmission [250].

In this chapter, we present an upgraded model of *TBMetapopPy* that incorporates a post-primary stage of disease in order to study how and where cavities form in the lung during infection. We first overview some important concepts for post-primary disease and cavitation and the structure of the human bronchial tree, then describe how the model has been amended to incorporate these factors into a new, more-granular environment and detail other dynamics changes. We outline an algorithm to

create a space-filling tree inside any 2-dimensional shape, and show how this was used to create an approximation of the bronchial tree for the second iteration of our *TBMetapopPy* model. We then present the results of experiments run using this iteration of the model, paying attention to the overall bacterial load of the system, as well as the size and location of lesions, showing that our model favours the apical localisation of larger lesions during the latent stage of disease, creating a preferential region just below the apex whereby conditions are ideal for bacterial growth. We incorporated a paradigm of post-primary disease, and show how a weakening of the immune system in the models leads to the apical lesions reaching bacterial loads seen in the initial lesion during primary infection.

5.2 Methods

5.2.1 Background

The human lungs are supplied by a single branching structure, the tracheobronchial tree (for a comprehensive review of the structure of the human tracheobronchial tree, see [191]). Air breathed in passes down the trachea, which splits into two main bronchi, one for each lung. This split is dichotomous, and almost all subsequent divisions in the bronchi are also dichotomous, although an occasional trichotomous split may be witnessed [112]. The main bronchi then divide into lobar bronchi, with each one supplying a lobe of the lung. The right lung contains 3 lobes, and the main bronchus divides once to supply the upper lobe, then again to supply the middle and lower lobes. The left lung has 2 lobes (upper and lower), and the left main bronchus divides in two to supply each. These lobar bronchi then divide to supply the various bronchopulmonary segment (as described in 4.2.1).

Subsequently, the airways divide, with about 20 to 28 divisions occurring[112, 166]. As the branches divide, the branching angle, diameter and length of child branches decreases [112]. Thus the bronchial tree is a self-similar structure, a fractal, with divisions in lower branches reflecting

divisions further up the tree [149], and the bronchial tree can be modelled using fractal techniques [91, 177]. Computational models of the human airway have been previously created using these fractal methods and fit to real-world data [60, 203].

At the termini of these branches are the terminal bronchioles, which are the last branches in the conducting zone. There are estimated to be approximately 25,000 terminal bronchioles in a single lung [113]. These then branch into respiratory bronchioles, and the surfaces of these respiratory bronchioles are interrupted by alveoli, where oxygen exchange and TB disease occurs.

5.2.2 Aims

In order to create a model that could better inform us of the specific areas within the lung affected during different stages of disease, a number of further developments were made to the initial *TBMetapopPy* model. This demonstrates the iterative nature of mechanistic model development as described in section 1.1.3 of this thesis: an initial model is made that conforms to our knowledge of the real-world dynamics, and this model is then refined as either our knowledge changes or we wish to explore the subject matter in more detail. In this section, we outline the various changes made to create the next version of the *TBMetapopPy* model.

The model presented in chapter 4 can be seen as model of primary and latent infection: an initial infection is seeded in the lower regions of the lung and the presence of this triggers and adaptive immune response. The increased immune activity brings the infection level down and contains it (but does not eradicate it), such that the disease never reaches the bacterial levels during the initial infection (albeit with oscillating bacterial levels).

Symbol	Parameter	Description
P	Perimeter	A series of (x,y) coordinates composing the external perimeter of the model environment
a	Branching start point	An (x,y) point, on P, where the branching process originates from
f	Branch length divisor	The length of the new branch as a fraction of the line dividing the perimeter in two
z	Minimum area	The minimum area needed to stop the branching process
S_V	Ventilation skew	The scale of differential of V
S_Q	Perfusion skew	The scale of differential of Q

Table 5.1: New required environmental parameters for the second iteration of *TBMetapopPy*

5.2.3 Environmental changes

The largest changes to the model were made to alter the underlying environment of the system within the `PulmonaryEnvironment` class. Rather than a hard-coded network as had previously been used, the model dynamically constructs the network based on parameters that are provided at the start of simulation. The new required parameters are listed in Table 5.1.

The process creates a 2D space-filling tree. A perimeter, P, is provided which forms the 2D reference space for the model. A single point on this perimeter, a, is also provided and the branching process begins at this point. An recursive process then occurs, an example of which is shown in Figure 5.1, which proceeds as follows.

a) An initial perimeter, P^*, for a shape is provided, as is a starting point, a^* on the perimeter.

b) A second point, $b**$, on P^* is chosen such that the line between a^* and b^* divides the shape created by P^* exactly in two[1]. The length, a^*b^* is calculated.

[1]All polygon areas are calculated using a Python implementation of the Shoelace formula, whereby a shape of co-ordinates [(x_1,y_1), (x_2,y_2),...,(x_n,y_n)] has area
A = $\frac{1}{2}|x_1y_2 - x_2y_1 + x_2y_3 - x_3y_2 + ... + x_{n-1}y_n - x_ny_{n-1} + x_ny_1 - x_1y_n|$

c) A point, c^*, on the line is calculated such that the length of the line from a^* to c^* is $\frac{a^*b^*}{f}$

d) A branch is created from a^* to c^*. The resulting shapes formed by bisection of P^* by the line from a^* to b^* then form the areas for the next child branches. The process continues recursively from a) using these two shapes.

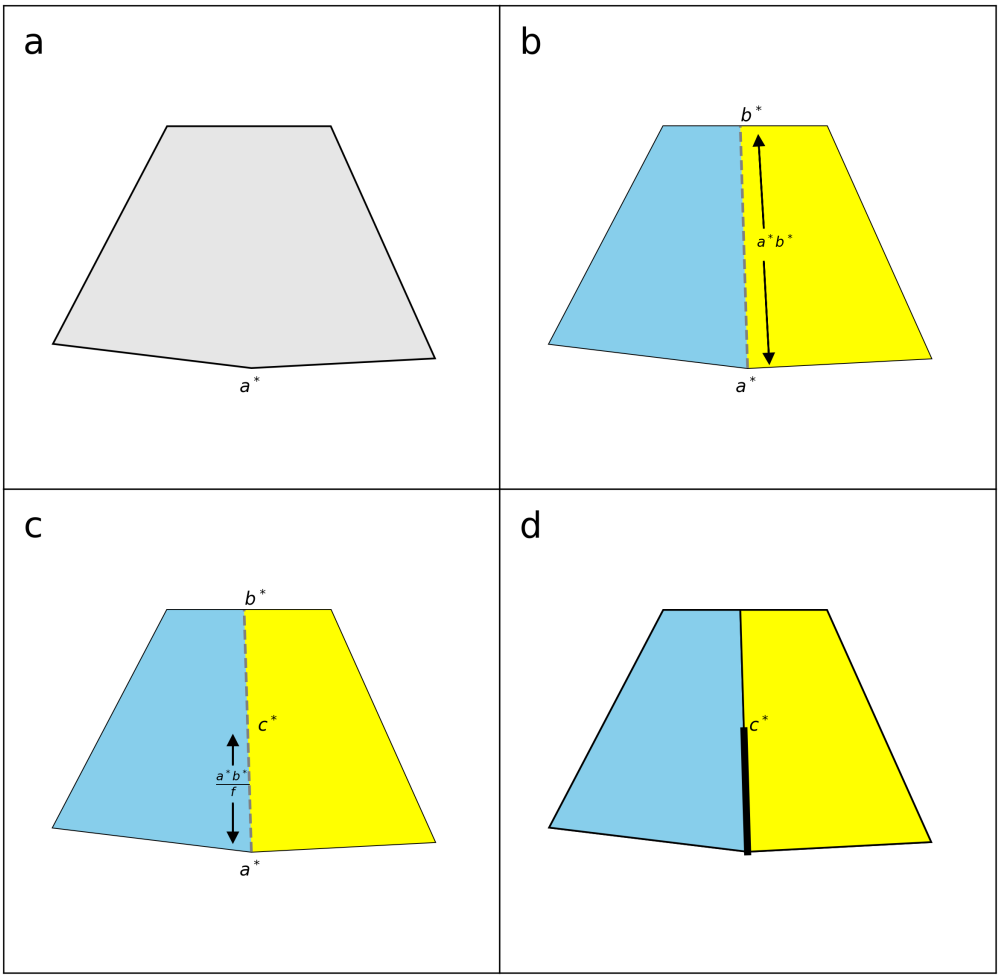


Figure 5.1: Example of the branching algorithm used to generate the environment within the second *TBMetapopPy* model.

Once the area of the child shapes drops below the threshold for the minimum area, z, then the branching process terminates. At this point, a patch is placed at the location that the final branch terminates. Thus

an abstraction of the bronchial tree is created, with a branching structure that fills the available space and ultimately terminates at each patch of the network, with each patch represents all the alveolar tissue that would be present at the end of that particular branch.

The values for environmental variables were also modified. Rather than using attributes set to arbitrary values as had occurred in the original model, the values were chosen to more accurately reflect the vertical position in the lung. It was decided that all V and Q values should sum to 1, thus meaning that the V and Q values at each patch represent the proportion of ventilation and blood perfusion from the whole lung that is directed to that patch. This simplifies the calculations for immune cell recruitment: the recruitment rate can be set to reflect the total numbers of cells found in the human lung (which data exists for) and the V and Q values then scale the number of cells at a given patch based on the number of patches and their vertical position. In order to achieve this, two new parameters, S_V and S_Q, were introduced to skew the differentials of V and Q respectively. Once the network is constructed, scales are created such a patch at very apex of the lung would have both V and Q values set to 1, whilst a patch at the very base of the lung would have a V value of S_V and Q value of S_Q. Patches are then assigned values according to these scales and based on their vertical position, as shown in equations 5.1 and 5.2, where y is the vertical position of the patch, y_{max} is y-coordinate of the highest vertical point of the area and y_{min} is the y-coordinate of the lowest vertical point.

$$V = 1 + (y_{max} - y) * \frac{S_V - 1}{y_{max} - y_{min}} \tag{5.1}$$

$$Q = 1 + (y_{max} - y) * \frac{S_Q - 1}{y_{max} - y_{min}} \tag{5.2}$$

Once every patch in the lung has been assigned these values, the sums of all values for V and Q are calculated and each patch's values are divided by these totals in order to normalise them and make them all sum to 1.

5.2.4 Event dynamics changes

Further changes were made to dynamics of the events within the *TBMetapopPy* model, to either conform to the new network dynamics, to improve the performance or to better match the known pathophysiology of TB. These changes are detailed below:

- In order to simplify the model, contact between members of the model was standardised, whereby each event that involves contact between members uses a half-sat formulation for its state variable. Furthermore, the dynamics of contact between bacteria and immune cells was altered: whereas in the previous model there existed two events for each extracellular bacteria and macrophage combination (one for bacterial destruction and one for macrophage infection), the second model uses a single event: bacterial ingestion. We assume macrophages ingest any bacteria they come into contact with. There is then a probability associated with the chance of infection: a macrophage internally destroy the bacteria or may fail to do so and thus become infected. This allows us to include a saturation dynamics for bacterial destruction: the destruction of bacteria by macrophages is reduced when the numbers are very high as crowding effects within the lesion prevent physical access to bacteria.

- Bacterial dissemination from the lymph into the lung is amended. In the new model, we assume that bacteria in the lymph will be constrained from dissemination as a lesion forms: the granuloma is intended to physically 'wall-off' bacteria and stop the spread to other regions. In order to model this, we use a new compartment, caseum , as an indicator of the development of a lesion - the greater the caseum, the less likely a bacteria will be to escape the lymphatics and disseminate elsewhere. In our model, the caesium compartment increases when an infected macrophage dies - either naturally or through bacterial mechanisms or immune mechanisms. Thus, there is a brief window after the bacteria have entered the lymphatics but before a solid lesion forms there, wherein bacteria can disseminate

and form lesions elsewhere.

- T cell dissemination from the lymphatics was amended. In the original model, all activated T cells disseminated from the lymphatics to the lung to fight infection. In the new model, not all areas of the lung become infected, and a more nuanced dynamics was needed to allow some T cells to remain in the lymphatics in order to fight infection there. Chemotaxis of T cells to specific areas through the detection of chemokines is well-known [9, 135, 155], and we represent this by directing migration of T cells based on M_I numbers: a "chemokine" attribute, K, is added to each edge, which is updated when the number of M_I in the adjacent lung patch changes. Migration of T cells then follows Equation 5.3, where K_L is the sum of all K values on the edges from the lymphatics to the lung (i.e. the total amount of infection in the lung) and M_I is the number of infected macrophages within the lymph. $\frac{{K_L}^{\sigma_T}}{{K_L}^{\sigma_T}+{M_I}^{\sigma_T}}$ is a scaling term that will have values between 0 and 1: as K_L gets larger compared to M_I (i.e. the infection in the lung outweighs the infection in the lymph), the term tends to 1 and the rate increases. As M_I outweighs K_L (i.e. the infection in the lymphatics outweighs the infection in the lung), the term tends to zero and thus the rate of T cell migration decreases. A power term, σ_T, allows us to control how the change in infection affects the change in T cell translocation - if σ_T tends to infinity, the rate becomes a simple switch: if M_I exceeds K_L, the term is zero and no translocation occurs, whilst if M_I is less than K_L, the term equals one and all T cells migrate.

$$T_A * \frac{{K_L}^{\sigma_T}}{{K_L}^{\sigma_T} + {M_I}^{\sigma_T}} \tag{5.3}$$

- Greater T cells numbers were required to contribute to the larger number of lesions present in this model. We therefore implement a mechanism that was not included in the previous iteration of the model: T cell replication. Once activated, T cells can proliferate into many clones [9], increasing the availability of T cells. This only occurs shortly after activation: to model this, we allow T_A members

to create new T_A members, but restrict this event to only occur within the lymph patches (at rate λ_T).

5.3 Results

5.3.1 Environment

For our simulations, we created a simple lung environment: a 2D rectangle whose height is twice its width: this is a rough approximation of the dimensions seen in human lungs [134]. Thus, we initialise the model with P = [(0,0), (0,100), (50,100), (50,0)], and we set the initial branching point a = (50,50), a point mid-way up the right edge of the rectangle, approximating the entry of the main bronchus into the right lung. We set the minimum area z = 0.05 and length divisor f = 2. Using these values, the algorithm creates a H-tree-like structure, with 65,536 terminal nodes.

For the environmental variables, we use S_V = 2 and S_Q = 3 [237].

5.3.2 Experiment

The second iteration of the *MetapopPy* model is computationally intensive - the larger number of nodes combined with the increased, more accurate, number of immune cells being modelled results in a greater number of computations per experiment. Therefore, it was necessary to delegate the computation of the simulation repetitions to a dedicated computer server with 8 cores. As *MetapopPy* was built upon the *epyc* library, we were able to use the `ClusterLab` class of that library in conjunction with an *ipyparallel* computer cluster running on the 8-core machine to improve the speed at which results were obtained from each simulation run.

We ran 32 simulations in total using the parameters as specified in the tables in Appendix A.

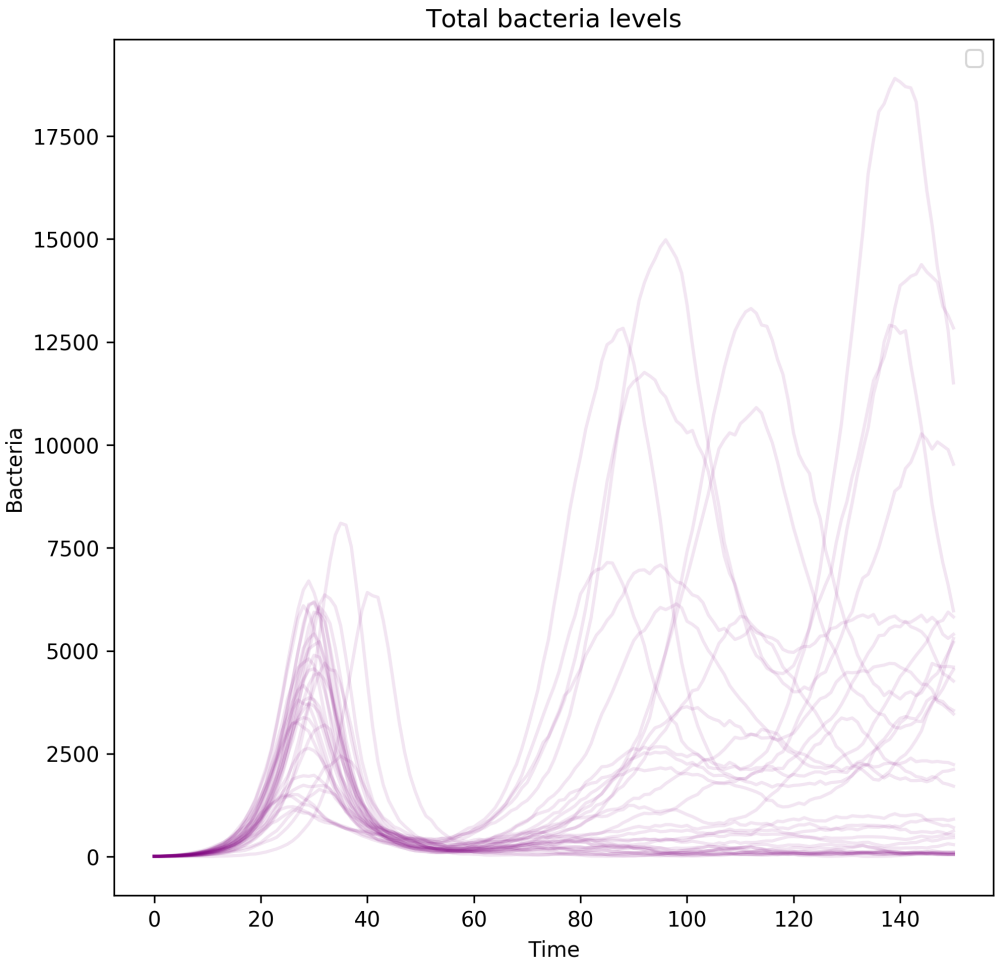


Figure 5.2: Total bacterial load of all patches over time within the *TBMetapopPy* system, showing individual counts for each of 32 repetitions

5.3.3 Total bacterial load and lesion size

We first examined the total bacterial load of the system over time. Figure 5.2 shows the results of the total bacterial load for each of the 32 repetitions over time. In the initial phase, a primary infection begins and reaches its peak between 25 and 45 days in all simulations. This is brought under control by the introduction of the adaptive immune response with low bacterial levels in all simulations by about day 55.

Whilst all simulations are relatively consistent to the point where the infection has been brought under control, with only minor variations in

the height and time of the peak of initial infection, the results vary much greater in the response after this point. Whilst many simulations tend to keep bacteria numbers at a low level, for those that do not, the total number present has a large amount of variation. But it is impossible from this Figure to accurately understand the whole situation: a high bacterial load may be a result of a single large lesion, or multiple small lesions. It is important to make the distinction between these two scenarios: multiple small lesions can be interpreted as being stable: the immune system is able to keep these lesions at a low level and stop bacterial growth; whereas a large lesion can be interpreted as a failure of the immune system to control bacterial growth at that location: this may lead to tissue damage, and then cavitation. Therefore, it is preferable for the host to keep the lesion size small, and it is important for us to understand the size of the lesions created.

We investigated the notion of bacteria distribution by tracking the average size of all lesions for each simulation, with the results shown in In Figure 5.3. Here, we see that the average lesion size after day 60 is much lower: indicating that while there are a high number of bacteria, they are often being contained in multiple smaller lesions than fewer, tissue-damaging large lesions.

5.3.4 Spatial distribution of lesions and lesion size

We then explored which spatial areas of the lung were being affected by the disease over the course of the infection, by looking at the average lesion size for each horizontal slice of alveolar tissue (i.e. we group patches together that fall within the same given vertical position boundaries). Figure 5.4 shows the correlation of average lesion size to vertical position in the lung during the primary stages of infection. The initial lesion (in the very lowest region) grows and peaks at around 30 days - this was the largest average lesion size seen over the whole course of the simulation time-frame. By day 50, the lesion has stabilised but bacteria have spread to over regions of the lung. At this point, these lesions are all at a very small average size and we see almost complete heterogeneity in size

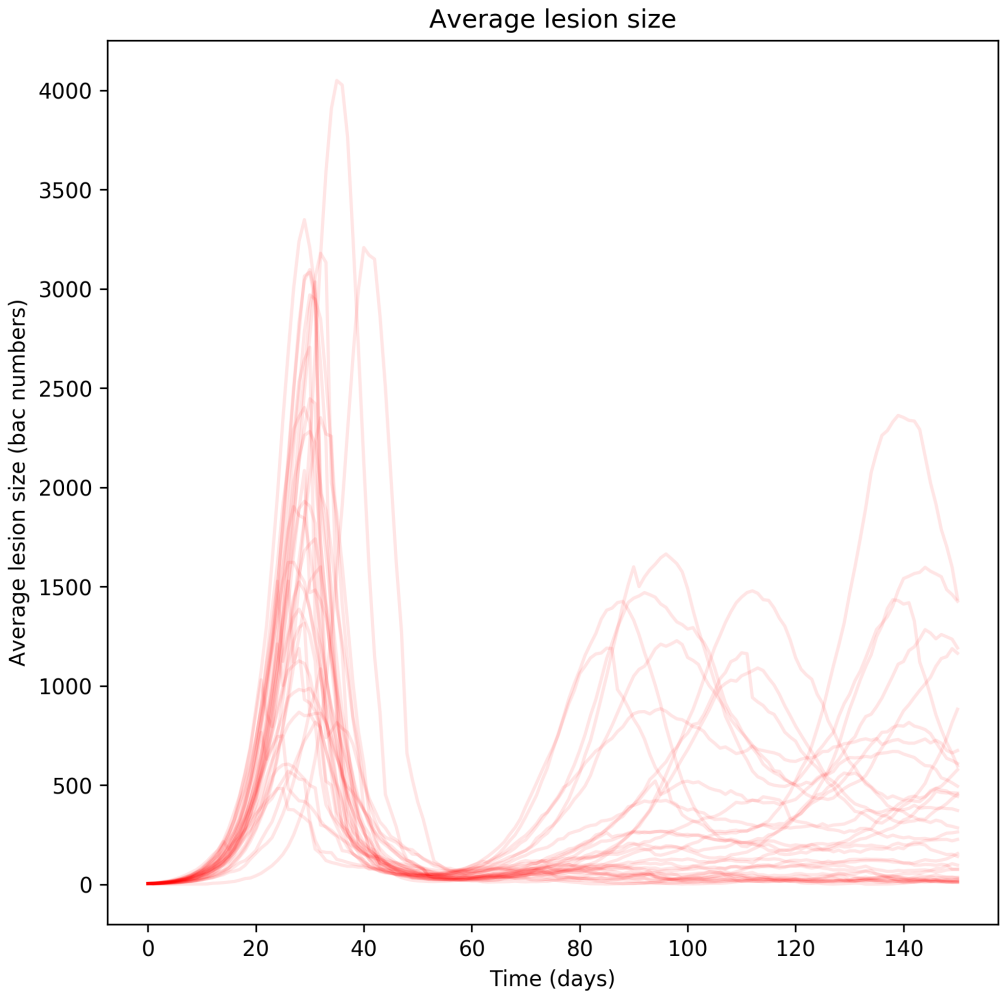


Figure 5.3: Average size (measured as number of bacteria) of all lesions over time within the *TBMetapopPy* system, showing individual averages for each of 32 repetitions

across the lung.

Figure 5.5 shows the average lesion size during the following stages of infection, from day 60 to day 150. We witnessed that whilst lesion starts homogeneous across the lung, as time progresses the lesions at the apical region start to expand at a greater rate than those towards the base. This is in stark contrast to the lesions during the primary stage, which are typically completely homogeneous in terms of bacterial load. The size of these lesions during latency fluctuates in a sinusoidal manner (see Figure

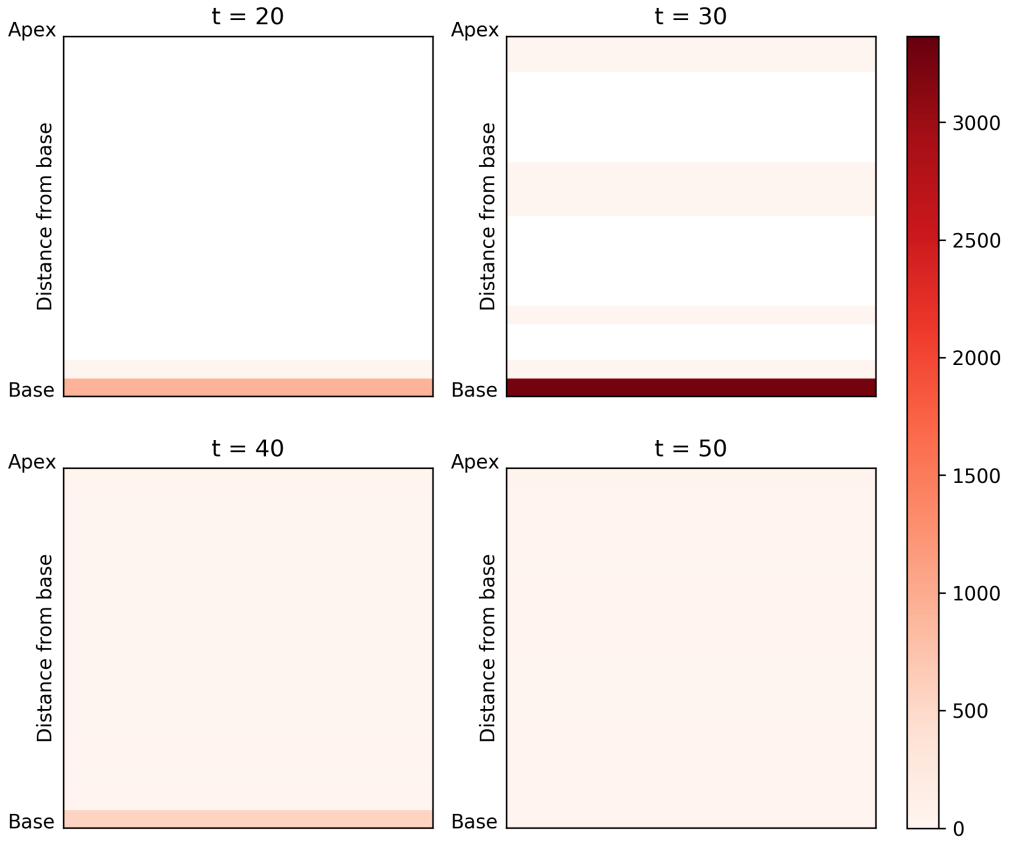


Figure 5.4: Vertical distribution of lesion sizes with *TBMetapopPy* simulations during the primary infection stage. Colour shows the average lesion size at each of 20 evenly sized horizontal slices of lung taken from 32 simulations.

5.3), but these average lesions sizes never reach the size seen during the primary infection. We further explored the numbers of lesions during infection, again stratifying the lesions based on their vertical position. Figure 5.6 shows the average number of lesions in each horizontal division of the lung per simulation. Here, we only counted patches where there was a bacterial presence - i.e. any lesions that were healed (no bacteria remaining) were not counted. In the primary stages, there is always a lesion present in the lowest basal region (the initial lesion where infection begins). By day 50, other lesions have begun to form across the lung, with a fairly even spread across the vertical dimension. Figure 5.7 shows the average number of lesions per horizontal section per simulation for later

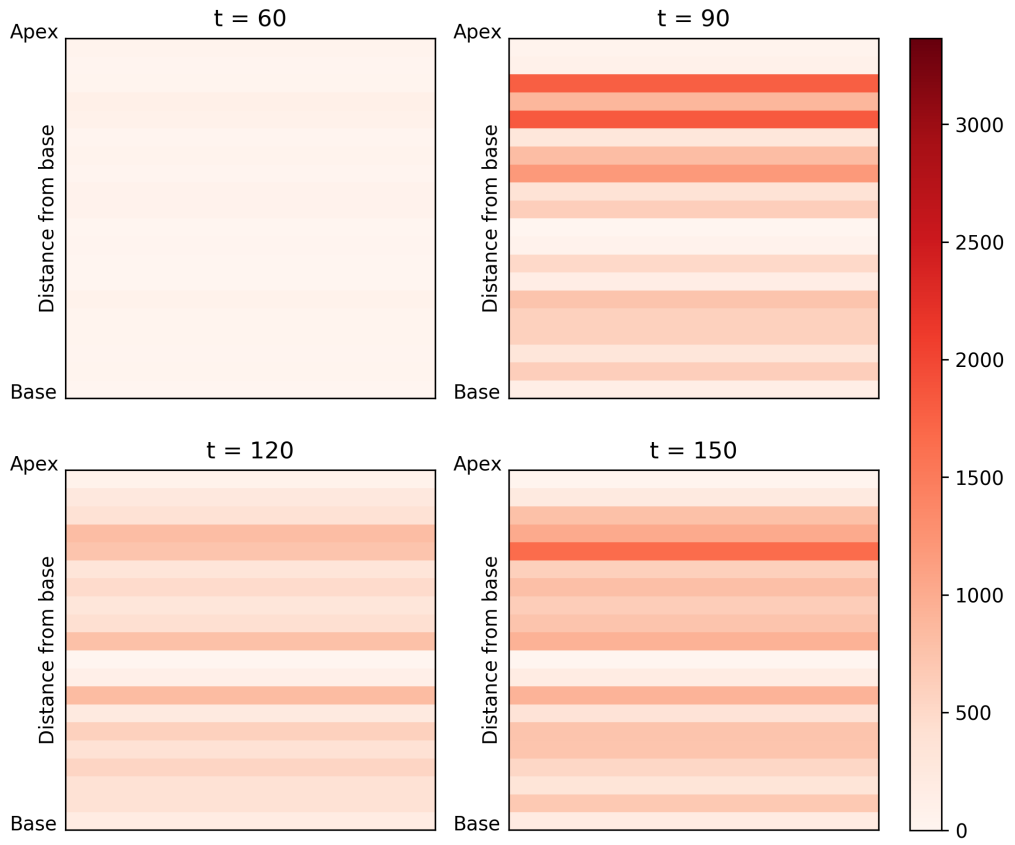


Figure 5.5: Vertical distribution of average lesion sizes with *TBMetapopPy* simulations during the latent infection stage

timesteps. At the very bottom section, the number of lesions decreases and reaches a level comparable to the rest of the lung. This occurs due to healing: the original lesion is, in many simulations, being completely healed of bacteria and thus is no longer counted. For the remainder of the lung, we witnessed some heterogeneity in the number of the lesions. By day 120, there appear to be two distinct regions with greater lesion counts: one at the very bottom of the lung, at one in the second quadrant from the top of the lung (marked A and B respectively on Figure 5.7).

5.3.5 Post-primary disease

Having established that our model permits a higher bacterial load towards the apices during latent infection, we then proceeded to model

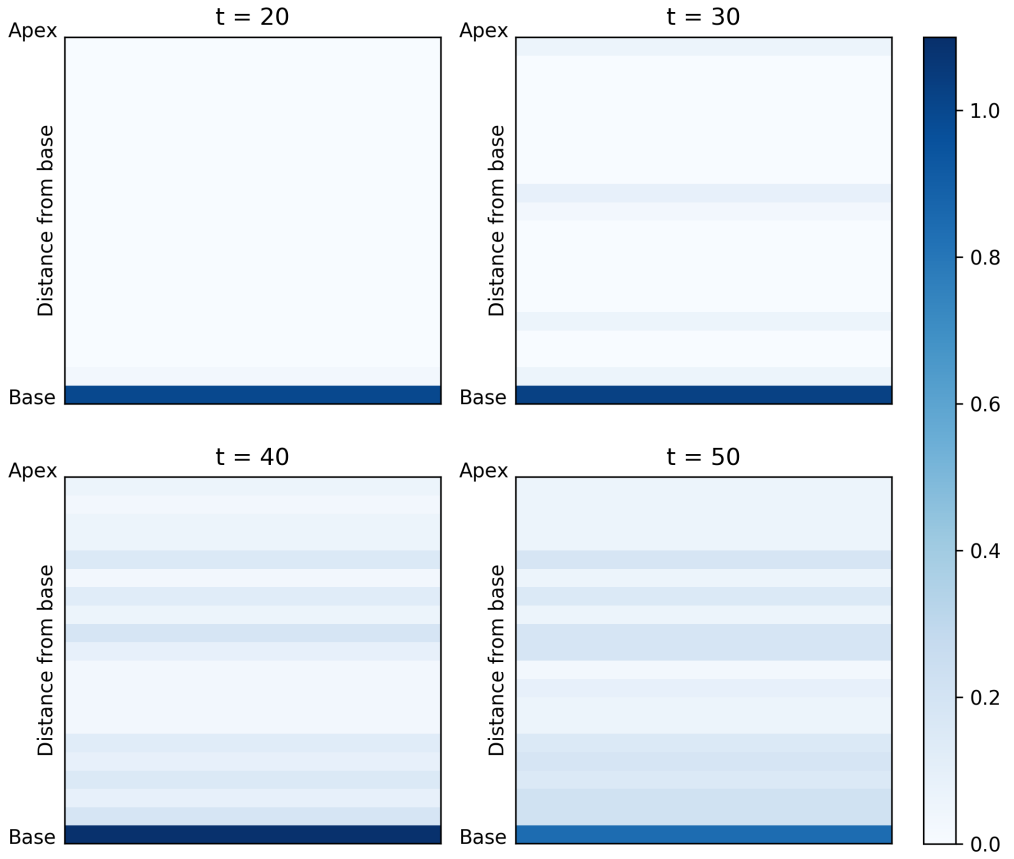


Figure 5.6: Vertical distribution of the number of lesions with *TBMetapopPy* simulations during the primary infection stage

the post-primary stage of the disease. As the actual causes of the change from latency to post-primary are unclear, and may not be the same for every individual case, we chose to model one case: a weakening of the immune system. In order to model this, we introduced a static event, which is set at the beginning of simulation and will always occur at the set time-point[2]. This event drops the T cell recruitment rate, α_T, by half, and is set to occur at 250 days, well after latency has been established. We again ran 32 simulations, this time running for 500 days.

[2]In order to implement this, we amended the event loop of *MetapopPy* to also check for static events. If a time-step, τ, is chosen that would increment the time beyond the next static event, the τ value is discarded and instead time is incremented to the time of the next static event. The loop then proceeds as normal. This may introduce a slight inaccuracy in the one event that would have been chosen.

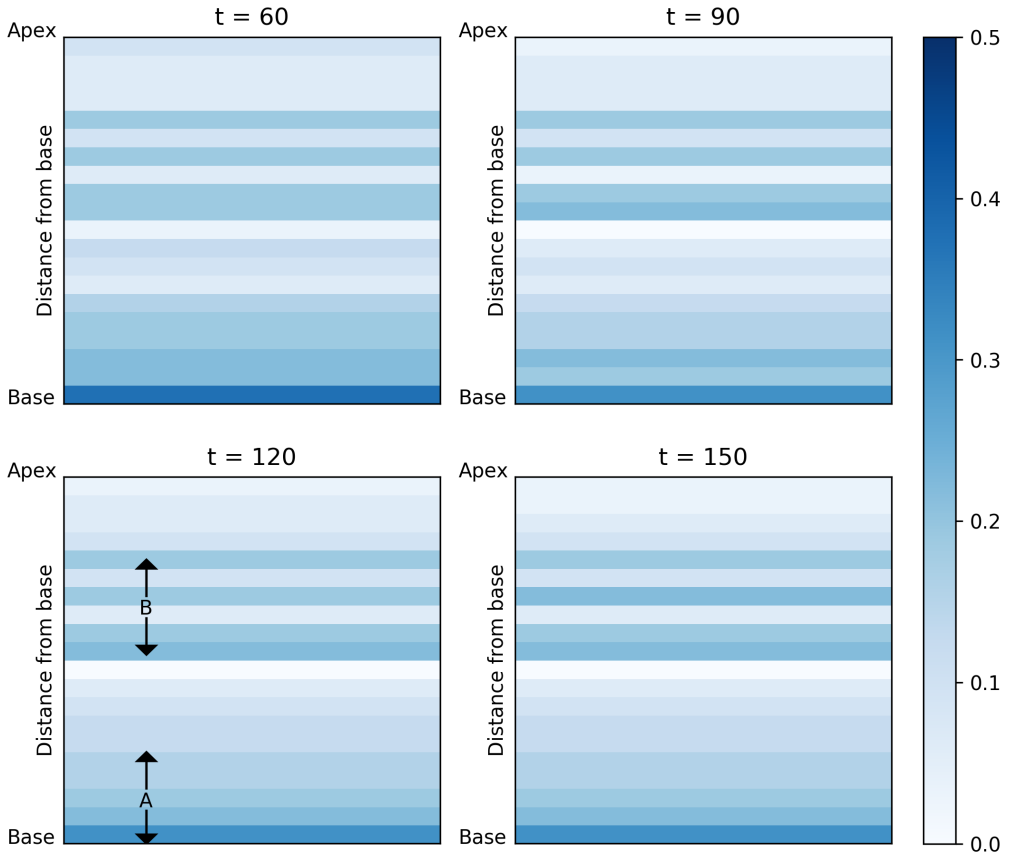


Figure 5.7: Vertical distribution of the number of lesions with *TBMetapopPy* simulations during the latent infection stage. Colour shows the average number of lesions present at each of 20 evenly sized horizontal slices of lung, with average taken from 32 simulations.

Figure 5.8 shows the distribution of the average lesion size for each of the 32 simulations. We see the same pattern as before for primary and latency disease. At the time of the T cell recruitment drop, the lesion sizes increase again, in many cases reaching levels similar to those seen at primary disease. We determine these to be a post-primary form of disease. We also note that not all replications reach this level: in some cases, where the average lesion size was low during latency, the average lesion size after the T cell recruitment drop increases but does not reach these high levels. In Figure 5.8 we show the spatial distribution of average lesion sizes within the lung. Again, the lesion size is large at base during primary infection (t=30), and this levels out at the end of

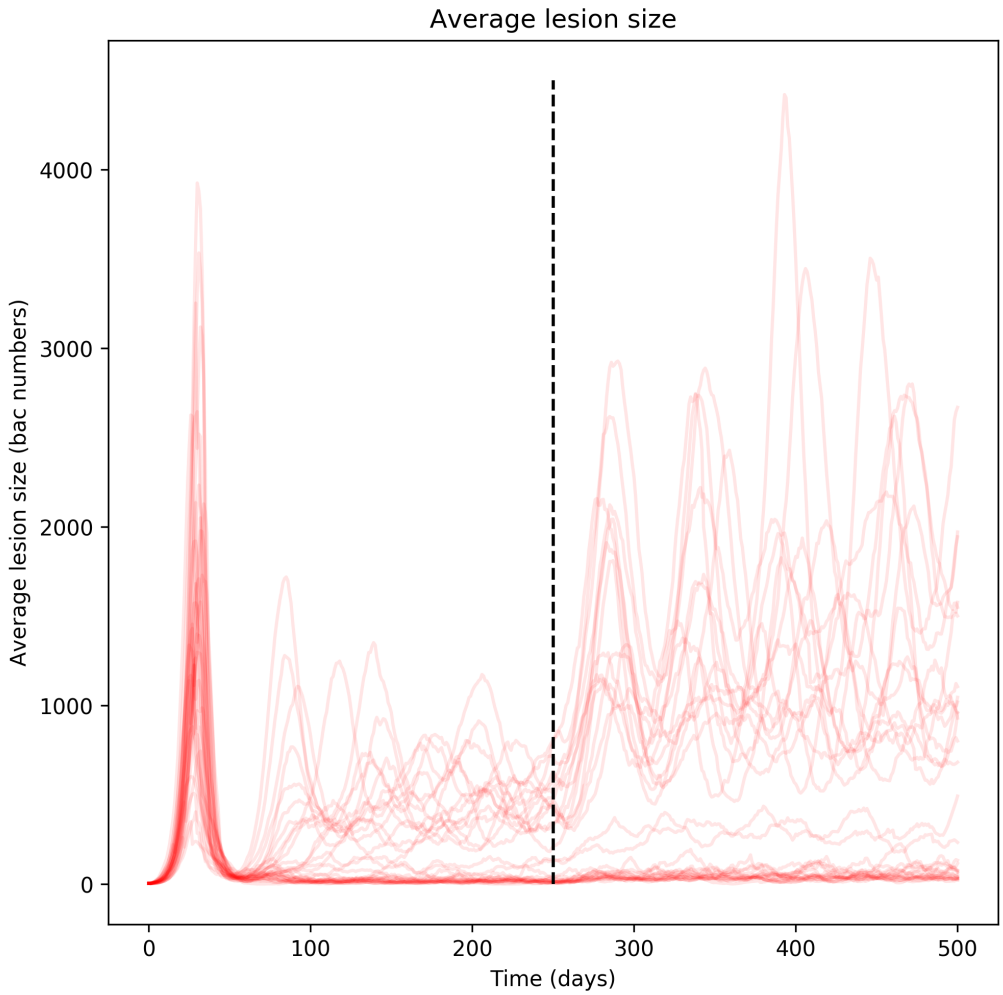


Figure 5.8: Average size (measured as number of bacteria) of all lesions over time within the *TBMetapopPy* system incorporating a T cell recruitment drop (black line), showing individual averages for each of 32 repetitions

primary infection and the beginning of the latent phase (t=60), with a slightly greater lesion size at the apex during latency (t=150). After the T cell recruitment drop, the average lesion sizes increase, with the largest, most damaging lesions appearing at the apex of the lung (t=400).

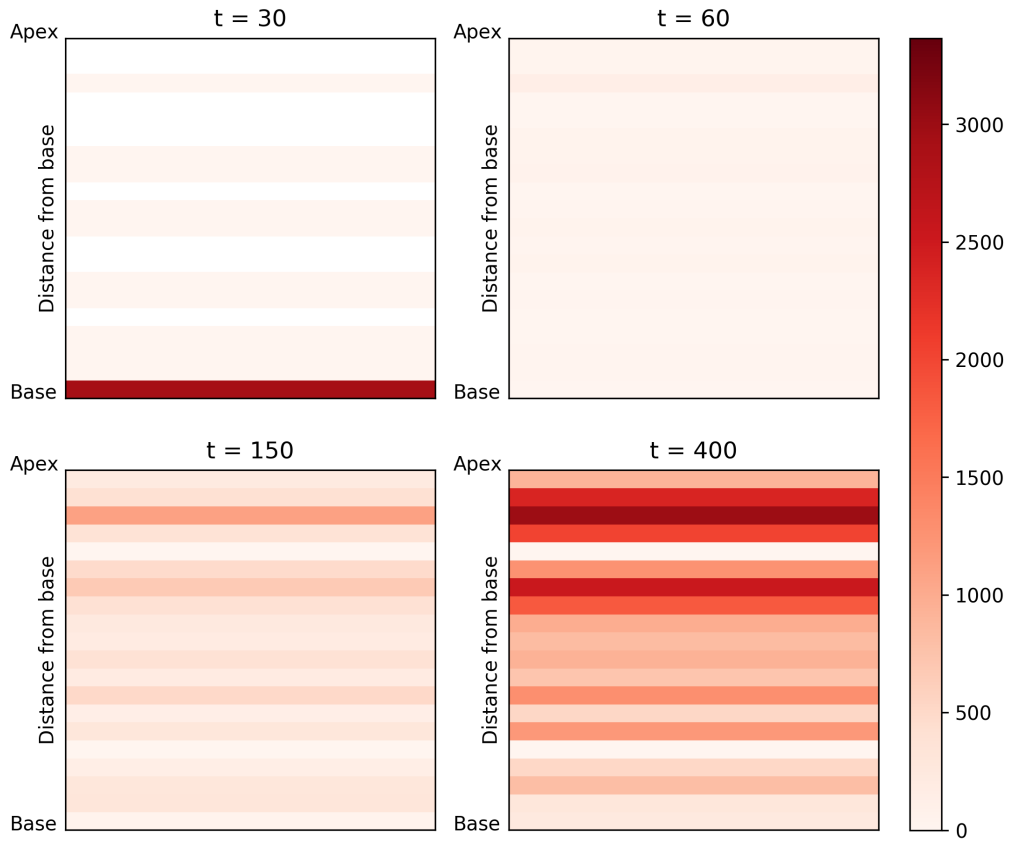


Figure 5.9: Vertical distribution of average lesion sizes with *TBMetapopPy* simulations during the primary (t=30), early latency (t=60), latency (t=150) and post-primary (t=400) infection stages.

5.4 Analysis

With this second iteration of the *TBMetapopPy* model, we are better able to understand the spread and size of lesions that occur during different stages of infection. Our findings show that after the establishment of a latent form of infection, the lesions that develop within the lung are heterogeneous in size. The conditions towards the apex favour bacterial growth, either through the availability of oxygen, the reduced immune response towards the apex or a combination of these and other environmental factors, resulting in larger sized lesions. This isn't a completely linear scale however, the areas with the largest sized lesion on average lie just below the apex (as seen in Figure 5.5). The very apex of the

lung is perhaps too poorly perfused to permit bacterial translocation in order to establish lesions. Therefore, there exists a location (about 6-10% of the total lung height from the apex) where the conditions are ideal from the point of view of the bacteria: the perfusion is just high enough to permit bacterial dissemination to the region but not so high as to allow a sufficient immune presence to keep bacteria levels very low. It is reasonable to assume that these higher bacterial levels at the apex during latent infection are what contribute to apical localisation during post-primary disease: they serve as larger reservoirs for bacterial growth if the immune system is compromised, and thus they are the areas we would most expect cavitation to occur as the immune system battles to contain a ever-growing infection. We would expect future iterations of the model to incorporate this concept of cavitation and the effect it would have on bacterial growth (e.g. providing more access to oxygen as cavities rupture into airways).

We also see a differential with the number of lesions (Figures 5.6 and 5.7). Once latency is established, the lesions predominantly form in 2 main regions: at the very base, and approximately 25% to 50% of the lung height from the apex. The reasons for this clustering are unclear: the lower region can be explained by the increased perfusion there, as more bacteria seed this region more lesions form there. The average number of lesions decreases moving up the lung, until a sudden increase signifying the start of the second region. It is possible that this also forms due to the perfusion values: as perfusion decreases up the lung the chance of bacterial seeding likewise decreases, by the immune response also decreases, meaning the likelihood of lesion sterilisation (removal of all bacteria) drops. This may reach a tipping point, where the reduced immune response becomes too low to sterilise lesions and thus more lesions occur as no sterilisation now occurs. These concepts could be explored by experimentation with a greater range of lung sizes and topologies.

A common occurrence in our findings (both in this model and the first iteration in Chapter 4) is an oscillating level of bacteria (as seen in the individual runs in Figure 5.2). Whilst there is evidence of some oscillating

bacterial levels during infection [198], as well oscillations in numbers of immune cells for animal models challenged with mycobacteria [221], the amplitude of these oscillations witnessed is much lower than those seen in our model.

We have created a full-lifecycle model in incorporating one possible mechanism of the reactivation of TB after latency is established: a weakening of the immune system. Our findings show that in such a scenario, the lesions throughout the lung will increase in size, with those at the top of the lung reaching levels similar to those present during the primary stage. If we interpret bacterial size as the main driving factor behind cavitation, then this would support the notion of the environmental heterogeneities within the lung being the driving force behind the apical localisation of post-primary disease.

5.4.1 Limitations

This second iteration of the *TBMetapopPy* model is an improvement on the first as it allows us to analyse in greater detail the spatial locations of bacteria within the lung, but it still has limitations as an effective simulation of TB within the human lungs. Both the environment and dynamics of the model can be modified to improve the model and we discuss both of these factors in turn here.

This iteration of the model simulates TB over a static, two-dimensional, rectangular-shaped lung topology. Whilst this is a useful abstraction for the purposes of demonstrating how environmental heterogeneity affects the disease pathophysiology, it is obviously simplistic compared to the real-world three dimensional anatomy of the human lungs. A three-dimensional could have profound impacts, particularly if a future iteration of the model were to also account for differential orientation of the subject: the environmental heterogeneities of the lung are believed to be caused by gravity, so changes in posture or orientation (e.g. lying down) would alter these environmental conditions and create preferential regions for bacteria in other dimensions. This could be explored if the

model was expanded to a third dimension.

In this iteration, we included environmental impacts upon TB but have not included the reverse: how the disease affects the environment. This occurs during TB through a variety of mechanisms, the clearest of which is cavitation: whereby the disease triggers erosion of the alveolar structure, allowing bacteria to access new regions of the bronchial tree. Incorporating this would require changes to both the dynamics and environments of the model: new events would be needed to cause cavitation to happen and these events would need to alter the structure of the network, possibly adding edges between patches.

Cavitation does not occur in every TB infected person, and its presence is known to be associated with higher bacterial loads in sputum [187], and thus understanding the causes of cavity formation is of great importance. Whilst the exact mechanisms by which this occurs remain controversial, it is believed that the cause is actually due to the host immune system, as evidenced by the fact that patients with HIV were shown to have less cavitary disease than those who were HIV-negative [28], and that treatment with immunosuppressive drugs prevented the formation of cavities in rabbits [248]. Therefore, using just bacteria numbers as a surrogate for cavitation may be too simplistic, and rather this may be a result of excessive levels of activated macrophages attempting to control larger bacteria numbers towards the apex. Future models should explore these aspects.

5.5 Conclusion

In this chapter, we have presented a second iteration of our *TBMetapopPy* model. This iteration expands on the original by introducing a more granular environment, with the environmental attributes attached to each patch now coming as direct result of the patch's vertical location within the lung. We have also extended our model to simulate the entire life-cycle of TB, introducing an external event that drives the system from

a latent form of disease into post-primary, reactivation, TB.

With our model, we were able to produce some of the significant phenomena seen during TB infection, particularly a greater lesion size towards the apex during latent and post-primary infection. But as well as reproducing these phenomena, it is important to be able to explain the causes of these as well: specifically, determining exactly which parameters within the model contribute to the behaviours we see in the outcomes of the model. In order to create better treatments for TB, we need to understand which environmental, immunological and bacterial factors are most influencing the bacterial growth, lesion formation and degradation and cavity formation. This concept is explored in the following chapter.

Chapter Six

Sensitivity and uncertainty analysis

In which we overview the role of uncertainty and sensitivity analysis in mechanistic model development, detail the creation of a framework to perform these analyses, and present our results of applying these analyses to our TBMetapopPy model.

6.1 Introduction

It is important to remember that models are abstractions of real-world scenarios and not exact replications. Every model requires some form of abstraction of the complete set of dynamics, and ideally, models should be as parsimonious as possible - only including the events that are absolute required to investigate the specific problem [150]. The more complex a model is, the more information needs to be provided to it, and with many real-world scenarios this information is very often not well-known [210]. This is particularly true with regards to TB research, where a lack of knowledge about the underlying dynamics occurring during infection leaves much uncertainty about the values of parameters used when modelling within-host TB dynamics. However, this complexity and the

uncertainty it generates is sometimes unavoidable if a useful model is to be created.

As seen in the previous chapter, the model we created to study TB across the whole lung is relatively complex and involves a large number of parameters, thus introducing greater uncertainty in the model output. Therefore, it is important to understand how uncertainty in the model parameters propagates through to impact the model outputs. In this chapter, we explore techniques to determine how uncertainty in a models output is generated from uncertainty in the model parameters, and methods apportion variance in the model output to specific individual parameters. We outline the development of a Python framework that allows these techniques to be applied to a model, which has been built to allow for the model simulations to be run concurrently across an *ipython* cluster, thus reducing computation time (which can be particularly heavy for these analyses as a wide parameter space is explored). We then use this framework to perform analyses on the model described in the previous chapter, with our results presented here. We show that perfusion is significant sensitivity for lesion size - indicating that this may be the most dominant environmental aspect that drives localisation towards the apex. We also show the importance of t-cells to the model, with many t-cell parameters having significant sensitivities.

6.2 Background

6.2.1 Definitions

In this chapter, we will review two important concepts, which we firstly define:

- Uncertainty Analysis (UA) is used to determine how uncertainty in the model predictions creates uncertainty in the model outputs [150]. It is important to note that at this stage, UA does not determine which of the uncertain inputs are more or less responsible - it is only concerned with the overall output uncertainty [210].

- Sensitivity Analysis (SA) intends to apportion the uncertainty in model outputs to uncertainty in individual model inputs [209].

Thus, UA should precede SA - uncertainty is estimated by UA and then apportioned to the inputs via SA (although this may not occur for cases of model calibration or optimisation where uncertainty does not need to be quantified [210]).

6.2.2 Uncertainty Analysis

6.2.2.1 Types and sources of uncertainty

Any model created is the combination of many parts, including but not limited to [210]:

- Mechanics - the coded events which define the interpretation of the real-world
- Model parameters - the parameters that define the rates at which events occur
- Initial conditions - the starting state of the model
- Resolution - the granularity at which a model operates

Each of these inputs can contain a level of uncertainty, and further uncertainty can be introduced through the computational costs (which can limit the number of simulations) and coding errors [85]. In practical terms, it is not always possible to vary all of these factors during UA (varying the mechanics would require the development of multiple different models which may be too time-consuming) and thus only a subset of these can be analysed, termed "input factors".

For mechanistic models, the input factors which we are able to analyse through UA typically consist of the model parameters and the initial conditions, which may be uncertain due to a variety of factors, such as natural variation (e.g. the size difference between human lungs in different people), errors in measurement or an inability to measure [150].

This type of uncertainty is termed 'epistemic uncertainty', where uncertainty of output arises from uncertainty of inputs. Conversely, 'aleatory uncertainty' is where the output of the model is uncertain due to inherent randomness within the model [108]. For deterministic models, all output uncertainty arises from epistemic uncertainty, whereas for stochastic models, both epistemic and aleatory contribute.

In order to perform UA, the k-dimensional parameter space needs to be explored, where k is the number of uncertain input factors. Exploring the full space can be impractical, due to the computational overheads of running the complete cross-product of all parameter values, or even impossible, as input ranges are often continuous and thus generate an infinite number of possible values. Therefore, sampling methods are needed to generate a finite number of samples which nevertheless sufficiently sample the entire parameter space.

6.2.2.2 Latin hypercube sampling

A variety of sampling methods exist, with one of the most commonly used being Latin Hypercube Sampling (LHS), a Monte Carlo technique favoured for its efficiency [107]. As with all Monte Carlo techniques, LHS requires each of the K uncertain inputs to define values in terms of a probability density function (pdf). Common choices for a pdf are a normal distribution (where evidence suggests a more common value for an input) or uniform distribution (where no *a priori* information on values exists) [150]. We then execute the following procedure:

1. We define a sample number, N_S. This will be the number of unique input parameter samples generated.

2. We stratify each input range into N_S parts, with each part containing an equal probability. For uniform distributions, this results in evenly sized stratifications; for normal distributions, the stratification size is smaller closer to the mean.

3. We construct an LHS matrix (of size N_S rows by K columns) by selecting one stratification (without replacement) at random for

each of the inputs. This forms the model input samples.

4. We use the LHS matrix to parameterise the model, with each row corresponding to the input values for a simulation. We run the model, generating an output matrix, Y, consisting of N_S rows [1].

6.2.3 Sensitivity analysis

Once the uncertainty within the model outputs has been established, the next step is to apportion this uncertainty to the uncertainty with the model inputs. This is the purpose of SA, and in performing this analysis, it allows a researcher to identify parameters and inputs that are critical to the model output [150], and conversely, to determine those inputs which have little impact on the output and thus can be 'fixed' to a single value [210].

SA techniques can be classified into two distinct groups [208, 211]:

- Local: these techniques involve determining the sensitivity of the model output to a single input. They involve fixing all inputs (except the input of interest) to nominal values, and varying the input of interest to determine the variation in the model outputs.
- Global: these techniques vary all inputs. They aim to determine the global influence of a factor on the model's output variance, including the effects of interactions between inputs [210].

Local techniques are very efficient - the number of input samples required is much lower than those needed for global techniques (especially when there are a large number of parameters) [211]. But local techniques are limited, only exploring a small amount of the k-dimensional parameter space, and are not appropriate for situations where there are a large number of uncertain parameters, or there are parameters with large ranges of values [150]. In these cases, global techniques should be used.

[1] The output matrix Y can be one dimensional when only one model output exists, or may have multiple columns for multiple outputs. Where multiple outputs occur, analysis should be performed on each output independently.

Two such examples of global techniques are discussed in the following text.

6.2.3.1 Partial rank correlation coefficient

Partial rank correlation coefficients (PRCCs) provide a measure of monotonicity between an output and an input after the removal of the linear effects of all inputs except the input of interest [150]. PRCC can be combined with LHS as follows [2]:

1. Parameter samples, X (of size N_S samples by K parameters), are generated and the model is run to create an output matrix, Y (of size N_S samples by R results), using the method of LHS described in section 6.2.2.2.

2. Each input X_i (i=1...K) in X is rank-transformed, as is each result Y_j (j=1...R) in Y, with the highest value given rank N_S, and the lowest given rank 1.

3. Linear regression models are fitted to each input ($\hat{x}_i$, equation 6.1) and each output ($\hat{y}_j$, equation 6.2).

$$\hat{x}_i = c_0 + \sum_{p=1, p\neq j}^{K} c_p x_p \tag{6.1}$$

$$\hat{y}_j = b_0 + \sum_{p=1, p\neq j}^{K} b_p x_p \tag{6.2}$$

4. The PRCC value between input x_i and output y_j is then calculated as the correlation co-efficient between the residuals (r_i = x_i - $\hat{x}_i$) and (q_j = y_j - $\hat{y}_j$), as per equation 6.3, where $\bar{r}_i$ and $\bar{q}_j$ are the respective means of r_i and q_j.

$$PRCC_{x_i,y_j} = \frac{\sum_{v=1}^{N_S}(r_{iv} - \bar{r}_i)(q_{jv} - \bar{q}_j)}{\sqrt{\sum_{v=1}^{N_S}(r_{iv} - \bar{r}_i)^2 \sum_{v=1}^{N_S}(q_{jv} - \bar{q}_j)^2}} \tag{6.3}$$

Statistical significance tests can then be performed to determine if a PRCC value is significantly different from zero, or whether two PRCC values are significantly different from each other [150].

[2] For an illustrated example of the process of an LHS-PRCC scheme, see [150].

6.2.3.2 Extended Fourier amplitude sensitivity test

In contrast to PRCC, the extended Fourier amplitude sensitivity test (eFAST) is a variance-based method [208], and is more suitable than PRCC for input-output trends that are non-linear and non-monotonic [150]. eFAST works by varying the parameters, causing variation in the outputs of the model, and then partitioning the variance, determining what fraction of the total output variance can be explained by individual inputs. The power of eFAST lies in how the inputs are varied: inputs are varied at different frequencies, and the strength of the input frequency on the output frequency is measured using Fourier analysis[3].

We now outline the procedure for performing eFAST. The original eFAST procedure was described in [208], and an amended procedure was detailed in [150]. The sampling process listed here is derived from both of these sources:

1. Each of the K uncertain inputs must be defined with a distribution type (e.g. uniform, normal, etc.) and appropriate values. For inputs where no *a priori* information exists, a uniform distribution is appropriate, whereas for parameters where a more frequent value is known, a normal distribution would be more suitable [150].

2. We add an additional 'dummy' parameter (the distribution here is irrelevant), giving a total of K+1 inputs.

3. In order to generate the input samples, we analyse each input, x_i, in turn.

4. We assign parameter x_i a high frequency, ω_i, as per equation 6.4, where N_S is the number of samples required per run and M is the number interference factor[4] (usually 4).

$$\omega_i = \frac{N_s - 1}{2M} \tag{6.4}$$

[3]Fourier analysis is used to study how a function can be approximated by a combination of simpler functions. A classic example of this is in sound waves - whereby an overall sound wave can be decomposed into its constituent pure frequencies

[4]This is the maximum number of Fourier coefficients that may be retained in calculating the partial variances without interferences between the assigned frequencies

5. We assign a low frequency to all other inputs. In order to do this, we calculate a maximum complimentary frequency, $\omega_{(-i)}$, as per equation 6.5.

$$\omega_{(-i)} = \frac{1/M}{\omega_i/2} \tag{6.5}$$

6. Frequencies are then chosen to satisfy the following conditions: a) the step between each frequency must be as large as possible and b) the number of inputs to which the same frequency is applied must be as low as possible.

7. Thus each parameter, x_q, has a frequency and a vector of values, g_q, in the range (0,1) are calculated for all parameters using the respective frequency, ω_q, as per equation 6.6, where s is a vector (of length N_S) that varies between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$, and ϕ_w is a random phase-shift in the range $[0, 2\pi)$[5].

$$g_q = \frac{1}{2} + \frac{1}{\pi} arcsin(sin\omega_w s + \phi_w) \tag{6.6}$$

   This function, defined in [208], creates an approximate uniform distribution of values between 0 and 1 across the range of s (see Figure 6.1). The values in g_q can then be converted into the appropriate values for the required distribution of x_q.

8. Each parameter therefore generates N_S values, and N_S samples are created. We then repeat this N_R times, with N_R representing the number of resamples. With each resample, we create a new phi_w for every parameter, thus exploring more of the parameter space.

9. We repeat steps 3 to 8, taking a different input as the input of interest.

10. Once all parameters have been processed, we pass the samples into the model to produce the required outputs. Therefore, the total number of samples is $N_S * N_R * (K + 1)$.

[5]The random phase-shift ϕ_w ensures that the search curve does not repeat itself and thus covers the whole parameter space.

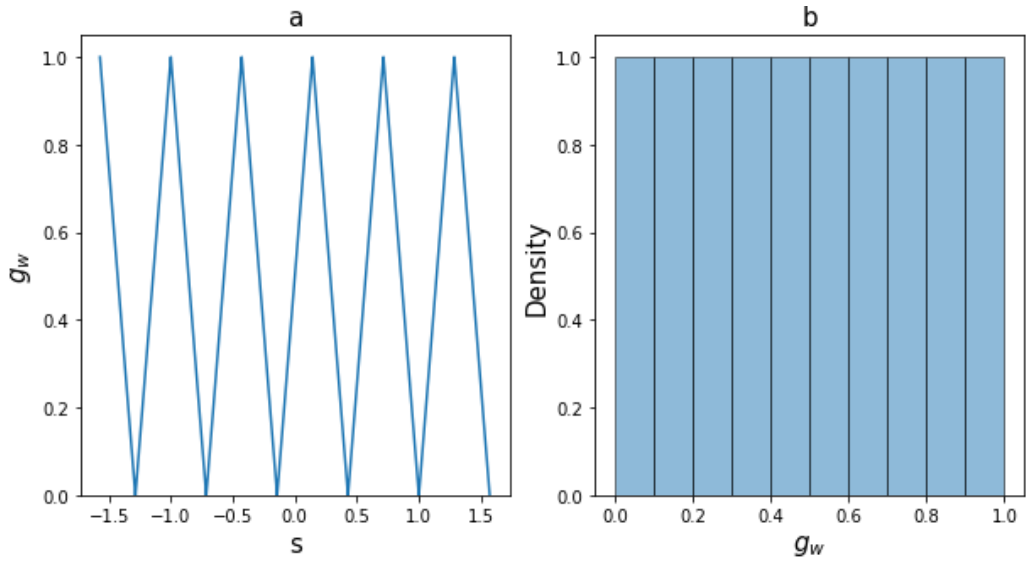


Figure 6.1: The sinusoidal function of eFAST. a) shows how as s varies between $\frac{-\pi}{2}$ and $\frac{\pi}{2}$, g_w values are generated between 0 and 1. These values are evenly distributed, as shown in b) which plots the histogram of density for values of g_w. Recreated from [208].

Having generated the full sample and produced the results, Y, these can then be analysed. This is done through Fourier analysis: for each input of interest and each resample curve, the output is Fourier transformed and the variance calculated from the Fourier coefficients at the frequency of interest. The first-order variance, S_i, represents the fraction of the variance of the model output that can be attributed to input x_i, whilst the total-order variance, S_{Ti}, includes the first-order variance and the variance due to higher-order interactions between inputs. eFAST produces non-zero sensitivity indices for parameters which the model is independent from: to counteract this, [150] proposed a scheme to introduce a negative-control dummy variable (step 2 above) and then compare input sensitivity indices against this known independent variable, to establish where the indices are significantly different from those of the dummy variable.

6.3 Methodology

In [150], the authors outline a methodology for performing UA and SA on a systems biology problem, with examples relating to within-host

modelling of TB dynamics, and we chose to follow this methodology due to its high relevance to our particular problem. Based one experiments with a variety of example models, the methodology described there suggests running both PRCC and eFAST analysis tests (as covered in the previous section) to understand model uncertainty and sensitivity, as each test measures a different output. Therefore, to perform UA and SA on our *TBMetapopPy* model, we implemented Python-based versions of both of these analyses.

One major issue, particularly with eFAST, is computation time: in order to fully explore the parameter space, a large number of samples will be needed, with this number increasing exponentially as the number of uncertain parameters grows. This is even more of a problem for a stochastic model: as samples from the same parameter set will generate different results, multiple repetitions need to be performed and averaged (as described in [150]) in order to reduce the influence of the aleatory uncertainty present in any model that contains random-based actions.

6.3.1 Existing software

In order to overcome the issue of high computational costs, we required a framework that could be run on a compute cluster. As we were unaware of any existing software that can perform UA and SA, we decided to develop our own framework, combining existing software to perform experiments on a cluster with existing software to generate relevant samples and perform analysis on the model results.

epyc, as previously described in Chapter 3, is a Python software library designed to handle experiment creation and running. The `Lab` class within *epyc* takes a series of parameters and possible values for each parameter. When an experiment is run in the lab, a parameter space is created based on the cross-product of all values of all parameters, and each parameter sample within the created space is sent to the model as the parameters with which the model is run. This process can be over-ridden, allowing for different user-defined parameter sampling mechanisms to be

implemented and create parameter spaces unique to the problem being researched. Results are saved within an instance of the `LabNotebook` class, and from here further analysis can be performed. Furthermore, *epyc* contains a subclass of the `Lab` class, `ClusterLab`, designed to run an experiment with multiple parameter samples over a defined compute cluster. These factors, combined with the fact that *TBMetapopPy* model is already an extension of the `Lab` class, made *epyc* an ideal starting point for our UA and SA framework.

The authors of [150] also provide MATLAB code with which to run PRCC and eFAST analyses[6] and we used this as reference guide. In Python, few libraries exist to perform UA and SA, with *SALib* [109] being one of the most prominent. This library provides the means to both create samples and analyse results for a variety of methods, including LHS and eFAST. The implementation of eFAST present in *SALib* is similar to that outlined in [150], but does not include the notion of resampling or of a dummy parameter. We chose to use the *SALib* implementation of eFAST, but amend them to conform with the [150] methodology. We also utilised the statistical methods of the *scipy* Python library, as this allowed us to calculate correlation coefficients and their associated p-values with ease.

6.3.2 Architecture

Our UA and SA framework subclasses two main classes of *epyc*:

- `ClusterLab`: to create parameter samples, and run each sample against the model, with multiple runs being distributed across a compute cluster.
- `JSONLabNotebook`: to store results and analyse them.[7]

Each of these classes was extended for each sensitivity test (PRCC and eFAST).

[6] http://malthus.micro.med.umich.edu/lab/usanalysis.html

[7] `JSONLabNotebook` is an extension of the `LabNotebook` class, where results are persisted within a JSON file. As the model would be run on a compute cluster on a remote machine, this was chosen in order to avoid any possible data loss during simulations.

6.3.3 Aggregating results

epyc allows for multiple repetitions to be run of every parameter sample created by the `Lab` instance. However, when recorded to an instance of the `LabNotebook` class, these values are stored in their full form - no aggregation of data across repetition occurs. In order to test for sensitivities (both PRCC and eFAST) across a stochastic model such as *TBMetapopPy*, averaging of results across multiple repetitions was required. Thus, a new class `AggregationNotebook` was created, as an extension of `JSONLabNotebook`, which contained functions to create a new data set from the experiment results, whereby results recorded across multiple repetitions were averaged together. This is then subclassed for specific sensitivity tests.

6.3.4 PRCC

For PRCC tests, the `ClusterLab` class was extended to create a new class of `LatinHypercubeLab`. Firstly, this class provides a function to add a parameter with a minimum value, a maximum value and a given number of stratifications. The parameter is added and the value range is calculated by dividing the range into the number of stratifications (based on the defined distribution). The `parameterSpace` function is overridden so that, rather than creating a cross-product of all parameter values, each uncertain parameter has its possible values shuffled (after a check to ensure all uncertain parameters have the same number of values/stratifications). Then, the first value is chosen from each shuffled parameter range and combined with the values for all certain parameters (those with only one value) to create a sample. Then the second value is chosen from each, and so on until all values have been chosen. Thus, the number of samples generated equals the number of stratifications specified.

In order to analyse results, the `AggregationNotebook` class is extended to create the new class of `LatinHypercubeJSONNotebook`, with a function `calculate_prcc`. This function takes a parameter and a result key, from which a PRCC value and a p-value are created. Firstly,

the aggregated parameter samples and aggregated results are ranked (using the *pandas* package). Then two linear regression models (equations 6.1 and 6.2) are fit (using the *sklearn* library): one between the parameter of interest and the remaining parameters, and on between the output of interest and the remaining parameters. The residuals of these are calculated, and then a Pearson correlation co-efficient is calculated from these two sets of residuals (using the `pearsonr` function of the `stats` module of *scipy* library). This returns two values: a PRCC value of the correlation and an associated p-value.

6.3.5 eFAST

For PRCC tests, the `ClusterLab` class was extended to create a new class of `EFASTClusterLab`. Similarly to `LatinHypercubeLab`, each uncertain parameter is defined with 3 values, one of which specifies the distribution type (normal/uniform etc.) and the other two to define the range (minimum and maximum or mean and standard deviation). The `parameterSpace` function is also overridden here. In this case, the uncertain parameters are determined and a new parameter, `dummy` is also added[8]. A frequency list (of length $K + 1$, K = number of uncertain parameters) is created, with the first value being the ω_i value defined in equation 6.4, and the remaining values are calculated as per steps 5 and 6 of the eFAST algorithm (section 6.2.3.2). A *numpy* array of length N_S is created, ranging from 0 to 2π.

Each parameter is processed in turn as the parameter of interest, and is assigned the ω_i value, with other parameters assigned values from $\omega_{(-i)}$. Then, for each resample (N_R), values between 0 and 1 are constructed for each parameter using the s array and the frequencies as per equation 6.6. Once these have been calculated (resulting in a k by $N_S N_R$ sized matrix), the 0-1 values are converted into values within the ranges of the parameters (using the `scale_samples` function from *SALib*).

[8] `dummy` is given a uniform distribution between 0 and 10, but the values are irrelevant as this parameter should have no bearing on the model result

Additional parameters are added to the parameter samples generated: these indicate which parameter is the parameter of interest and which resample number the sample belongs to and which run number it is within the resample. By default *epyc* shuffles the order in which samples are processed: however, the analysis of eFAST uses a Fourier transform and thus this requires the data to be ordered. Adding these additional parameters allows the data to be re-ordered later.

The `AggregationNotebook` class is extended to create the new class of `EFASTJSONNotebook`. M, N_R and N_S values are set during the running of experiments, and these are persisted within the JSON file. A `generate_sensitivity_indices` function returns values of S_i and S_Ti for each parameter and result combination. This function is adapted from the eFAST `analyze` function of *SALib*, and modified to accommodate the resampling introduced. Firstly, the aggregated data is re-ordered by parameter of interest, resample and run number (in that order). Then for each parameter of interest and each resample, the S_i and S_Ti are calculated as per the *SALib* function. This function returns a dictionary of parameter and results as keys, with each having a value consisting of two lists: one for S_i values (of length N_R) and one for S_Ti values (also length N_R). These values can be averaged to generate values for each combination of result and parameter.

6.4 Verification

Having developed this framework, it was necessary to run some verification tests to ensure that the output generated was correct. In order to do this, we attempted to duplicate some previously published results in SA. As the methodology present in [150] had been followed to create this framework, we chose the examples from within this work as a suitable means of comparison.

The papers defines a predator-prey (Lotka-Volterra) model which tracks two populations over time, prey ($Q_{LV}(t)$) and predators ($P_{LV}(t)$), using

Parameter	Mean	Standard deviation
α_{LV}	1.5	0.01
β_{LV}	1	0.2
σ_{LV}	3	0.2
δ_{LV}	1	0.01

Table 6.1: Parameter values used to verify the UA and SA framework. All parameters are normally distributed.

equations 6.7 and 6.8[9].

$$\frac{dQ_{LV}}{dt} = \alpha_{LV} Q_{LV} - \beta_{LV} Q_{LV} P_{LV} \tag{6.7}$$

$$\frac{dP_{LV}}{dt} = -\sigma_{LV} P_{LV} + \delta_{LV} Q_{LV} P_{LV} \tag{6.8}$$

We created a model (an extension of `epyc.Experiment`) and used the parameters values in Table 6.1 as the distributions for our framework of LHS and eFAST.

6.4.1 PRCC verification

Figure 6.2 shows the comparison of results between a) PRCC results generated from MATLAB code supplied with [150] and b) our UA and SA framework. It was necessary to use the MATLAB to generate PRCC values as we believe the paper contains an erratum: parameter α_{LV} is shown as having a positive correlation (and not being significant), whilst the actual code provided with the paper gives the opposite result: the parameter has a negative PRCC value and is significant. The significance factor is also contrary to the data supplied within the paper: Table 1 states α_{LV} is significant whilst the graph shows it is not. We see that the PRCC values generated by our code are very close to those generated from the MATLAB code (see Table 6.2 for results).

6.4.2 EFAST verification

Figure 6.3 shows the comparison of results between our framework (B) and the results presented in the paper (A). Whilst we see strong agree-

[9] In order to differentiate the parameters and populations used for verification from those used for our TB model, we subscript the verification components with $_{LV}$

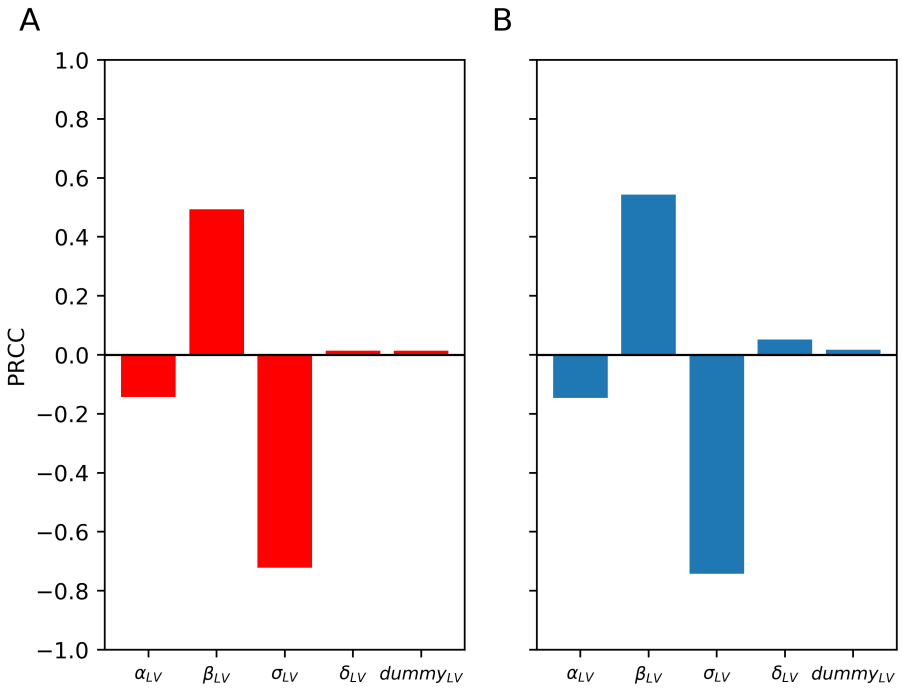


Figure 6.2: Comparison of PRCC results for a Lotka-Volterra model from a) [150] and b) our UA and SA framework.

Parameter	PRCC	p value	Significant?
α_{LV}	-0.14597746	3.564e-6	Yes
β_{LV}	0.544	5.750e-78	Yes
σ_{LV}	-0.742	1.378e-175	Yes
δ_{LV}	0.0516	0.102	No
dummy	0.0175	0.581	No

Table 6.2: Lotka-Volterra model PRCC values of parameters for output Q at time t=9, with associated p-values (significant when p<0.01).

ment between the two results in terms of the sensitivity indices created (for both first-order S_i and total-order S_T), there is some discrepancy with the standard deviations. The results in [150] have much higher standard deviations than the range produced through our repetitions for α_{LV}, δ_{LV} and the dummy parameter. The reasons for this are unclear: they may be due to errors in the published paper or because of the differences between the random number generator processes used in the original

Parameter	Index	Value	T-test (p value)	Significant?
α_{LV}	S_i	0.00276	0.897 (p=0.396)	No
α_{LV}	S_T	0.0684	-1.782 (p=0.113)	No
β_{LV}	S_i	0.175	100.865, (p=1.042e-13)	Yes
β_{LV}	S_T	0.630	68.377 (p=2.329e-12)	Yes
σ_{LV}	S_i	0.338	30.850, (p=1.325e-09)	Yes
σ_{LV}	S_T	0.805	72.932 (p=1.392e-12)	Yes
δ_{LV}	S_i	0.00281	0.998 (p=0.348)	No
δ_{LV}	S_T	0.0676	-2.003 (p=0.080)	No

Table 6.3: Lotka-Volterra model eFAST sensitivity indices of parameters for output Q_{LV} at time t=9, with significance checks of t-tests run between resamples of each parameter and compared to the dummy variable (significant when p<0.01).

paper and our Python code. These differences do not affect our significance checks in comparison to the Marino model: we find β_{LV} and σ_{LV} are significantly different from the dummy parameter for both first- and total-order indices (i.e. p<0.01) whilst α_{LV} and δ_{LV} are not significantly different for either (i.e. 0.01<=p) (see Table 6.3).

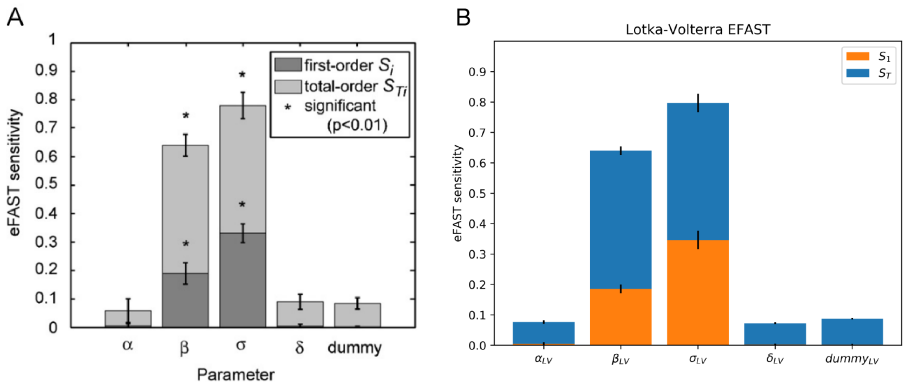


Figure 6.3: Comparison of eFAST results for a Lotka-Volterra model from a) [150] and b) our UA and SA framework. Error bars indicate 2 standard deviations above and below the mean.

6.5 Results

6.5.1 Parameters and outputs

In order to perform a comprehensive analysis of the second iteration of the *TBMetapopPy* model, we applied a range and/or distribution to every network-building parameter and event parameter (see Appendix B for the full list of parameter values and distributions used). Where an appropriate value could be derived from the literature or existing models, we applied a normal distribution using the derived value as the mean, with a reasonable standard deviation added (e.g. replication rate standard deviation values were chosen to ensure positive non-zero values were produced). For values where no existing value could be determined (i.e. rates and half-saturation values for contact between cells), uniform distributions were chosen, and given large ranges to allow for a wider range of possible values.

There exists a large number of possible outputs from the model: at each time-step, the entire state of the network is output and these values could be aggregated in great many possible ways to produce summary statistics. For the purposes of our SA, we examined three outputs, as listed in Table 6.4. These values vary over time, as the system moves between different stages of infection. Therefore, we track the SA values of each parameter and output combination over a time period, chosen as being day 1 of infection to day 120 (by which time the primary infection has been contained and a latent infection established).

We chose to only perform sensitivity analyses on the primary and latency stages of infection: our model of post-primary disease requires one of the parameters (α_T) to be changed during simulation. The sensitivity analyses we have outlined in this chapter are unable to account for changing parameter values so would not generate accurate results for a post-primary stage.

Output	Description
Ω_T	The total number of bacteria (of all states) present within the system
Ω_N	The total number of lesions present within the system (i.e. patches where bacteria count > 0, does not include lesions which have been sterilised)
Ω_S	The average size of all lesions within the system (i.e. the average number of bacteria per patch, includes patches where the lesion has been sterilised)

Table 6.4: Outputs from *TBMetapopPy* used for sensitivity analysis

6.5.2 PRCC

As per the recommended methodology published in [150], we began by first calculating PRCC values for all parameters against all output variables over the time-span of 120 days. We chose to use 100 stratifications of the input parameter ranges, and used the average results of 20 repetitions in order to reduce the contribution of aleatory uncertainty. As noted in section 6.3.1, the use of parallel computing reduces the computational time required for running this type of analysis: our simulations were run upon an 8-core machine.

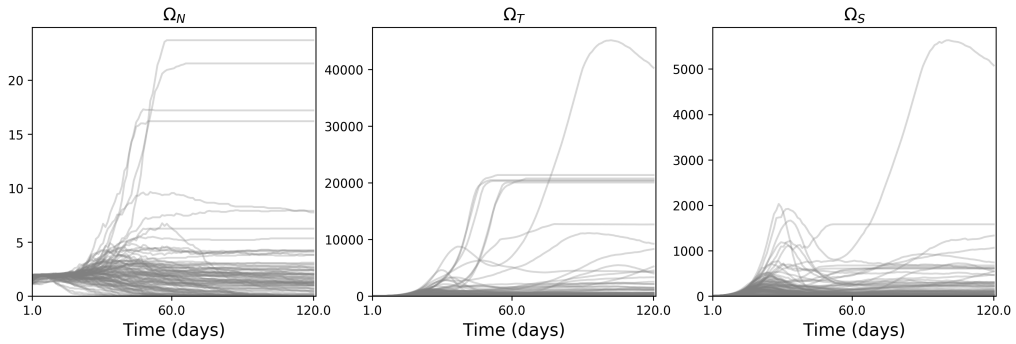


Figure 6.4: Uncertainty analysis plot of each output over time from each of the 100 averages (of 20 repetitions) for each parameter sample

We first plotted the results of our three outcomes in order to perform our

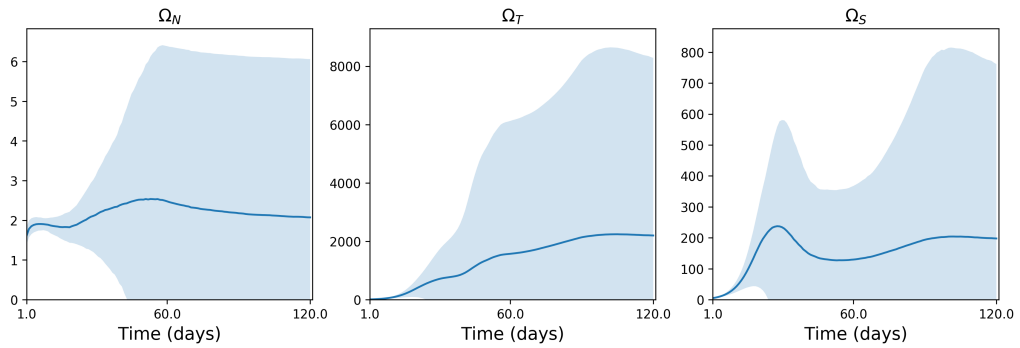


Figure 6.5: Uncertainty analysis summary statistics plot of all outputs over time from each of the 100 averages (of 20 repetitions) for each parameter sample.

UA and understand how uncertain our outcomes were. These are shown in Figure 6.4, which shows the plots of each of the 100 aggregations of 20 repetitions for each parameter sample generated by the LHS method, and summarised in Figure 6.5, which shows the mean and standard deviation over time of these outputs.

From these plots, we note that there is a high degree of uncertainty in our output variables, particularly in the later stages of infection (as seen by the large errors present for all three outputs in Figure 6.5. From Figure 6.4, we see the effects of aleatory and epistemic uncertainty: changes in the model parameters can lead to different scenarios in the outcomes (evidenced by the grouping of results at time t=120.0, particularly for Ω_S and Ω_T), while aleatory uncertainty creates minor fluctuations within these outcomes.

Having established the uncertainty within our model outcomes, we then generated PRCC values for each parameter and output combination over the time-scale of the simulations. The full set of plots and significant sensitivities are presented in Appendix C, and in this section we overview some of these results.

6.5.2.1 The influence of environmental parameters

The three environmental parameters, S_V, S_Q and S_G, all exhibit differing sensitivities upon the outputs (see Figure 6.6). For these variables, an increases indicates a increase in the heterogeneity within the lung; for example, a greater S_V value results in more ventilation being directed towards the base of the lung and less towards the apex, and similarly for S_Q and S_G. For the number of lesions, the environmental parameters show little sensitivity, with increased drainage at the base showing a slight negative PRCC value during the early stage of infection, implying better movement of immune cells to the lymphatics can reduce the number of lesions during the primary stage, but this reduces as latency is established. The sensitivity of the S_Q and S_V have little impact on the number of lesions throughout the infection.

After a period of around 30 days, the PRCC values for S_Q against

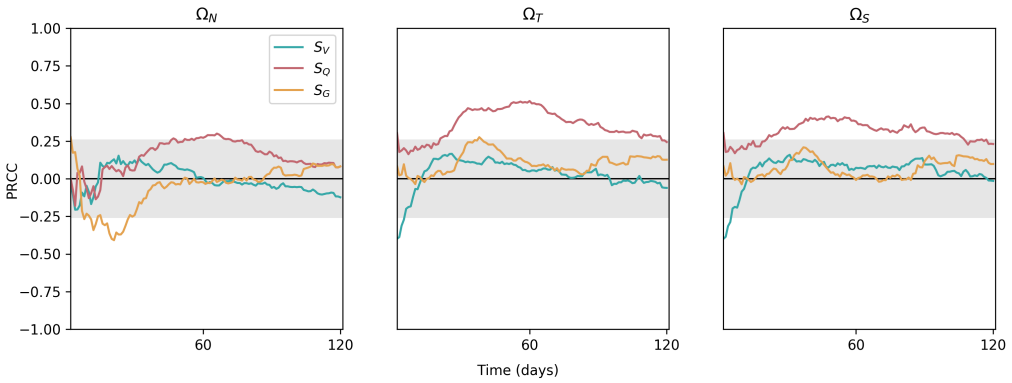


Figure 6.6: Sensitivity plots for environmental parameters within *TB-MetapopPy*, plotting the PRCC value of each parameter against one of the three outputs over time. Grey shaded area shows non-significance ($p<0.01$).

both the total number of bacteria and the average lesion size become significant. This implies that the differential in perfusion across the lung has little impact on the primary infection, but does impact the conditions present during latent infection, with a greater differential increasing both the total bacteria present within the patient and the size of the lesions they have developed.

6.5.2.2 The importance of T cells

Figure 6.7 shows a selection of the sensitivities of parameters related to T cells against the outputs of the model. We note that the model is highly sensitive to many of these parameters, indicating that T cells play a very important role in the course of infection. For the number of lesions, the sensitivities for most parameters remain in the insignificant range, but for most these values become significant after the primary stage of infection. T cell activation is very important, with ϵ_T having negative values (as greater rates of activation decrease the number of lesions) and $\theta_{\epsilon T}$ having negative values (as an increased half-sat reduces the rate of contact between T cells and APCs). This trend for T cell activation also occurs for the other two outputs. The speed at which T cells destroy macrophages and the contact rates allowing this to occur have high sensitivities for all three outputs, indicating this is an important process to maintain control over infection. Establishing a sufficient T cell presence is also an important factor, with T cell replication have a strongly negative PRCC value, as the proliferation of T cells is necessary in the model to fight bacterial growth.

6.5.2.3 Bacterial replication

Our PRCC results show contrasting findings for the replication of intracellular and extracellular bacteria on the outputs, as shown in Figure 6.8. Intracellular replication (λ_I) has a positive PRCC value for all three outputs during the initial primary infection (albeit with non-significance for Ω_N). This value becomes negative and remains non-significant until briefly becoming significant for all three outcomes at around 30 days. This is in contrast to the two extracellular replication rates, which at 60 days are both positive. This suggests there may be some benefit to bacteria by growing slower when intracellular, possibly due to slower growth reducing the rate of cell death of the macrophage, and thus keeping the bacteria in an environment where they cannot be harmed for a longer period of time.

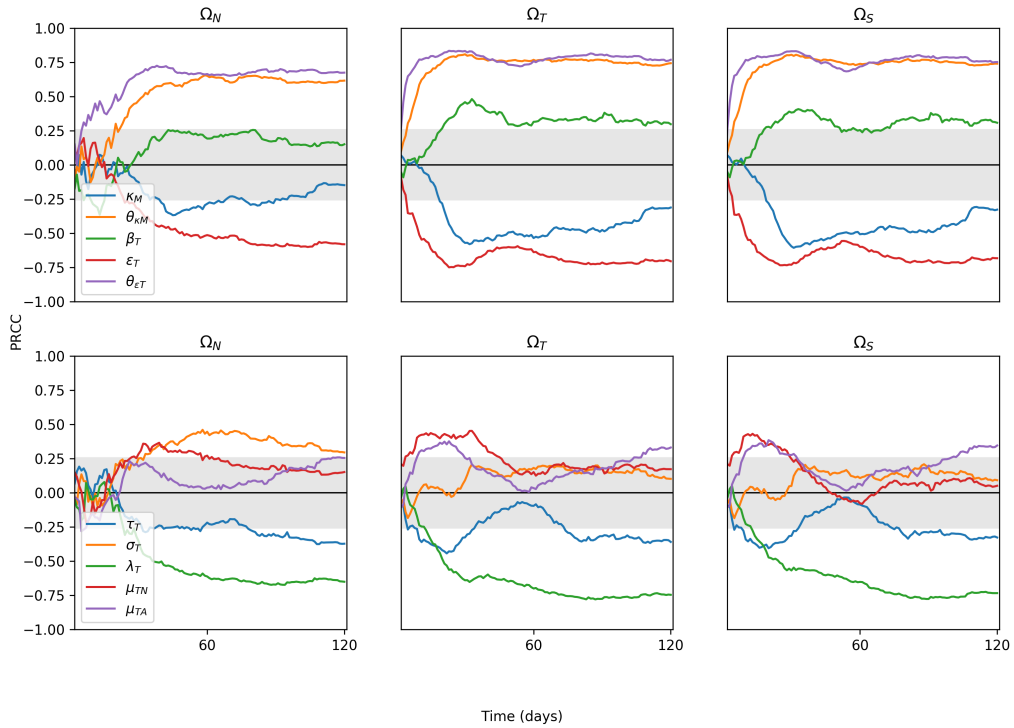


Figure 6.7: Sensitivity plots for T cell related parameters within *TBMetapopPy*, plotting the PRCC value of each parameter against one of the three outputs over time. Grey shaded area shows non-significance (p<0.01). Only parameters with sustained significant PRCC values against one output are displayed.

6.5.2.4 Translocation of macrophages is more important than translocation of dendritic cells

One perhaps surprising result is that the translocation of macrophages (τ_M) had higher PRCC value than translocation of dendritic cells (τ_D) for most of the time-frame of infection, as shown in Figure 6.9. τ_M occurs at a much lower rate than τ_D in our model (1e-4 - 1e-1 compared to 0.1 - 1.0 ranges respectively), indicating that this process is an important part bacterial containment within our model, possibly as the macrophages carry a larger quantity of bacteria than the dendritic cells, and thus draining more of them to the lymph nodes and exposing them to activated T cells faster is a suitable mechanism to improve the fight against the bacteria proliferation.

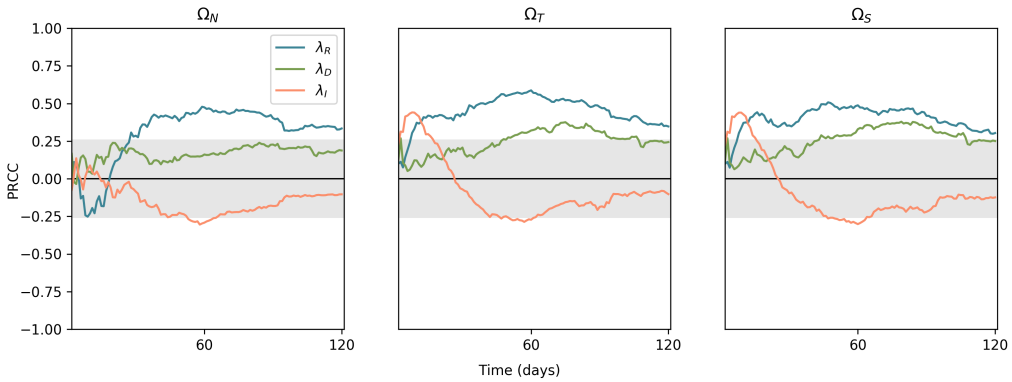


Figure 6.8: Sensitivity plots for bacterial replication parameters within *TBMetapopPy*, plotting the PRCC value of each parameter against one of the three outputs over time. Grey shaded area shows non-significance (p<0.01).

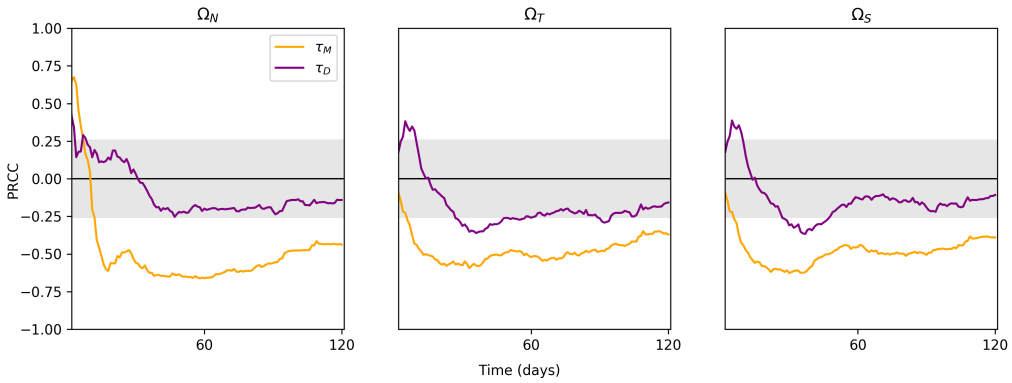


Figure 6.9: Sensitivity plots for immune cell migration parameters within *TBMetapopPy*, plotting the PRCC value of each parameter against one of the three outputs over time. Grey shaded area shows non-significance (p<0.01).

6.5.2.5 Parameters affecting different stages

As noted, many parameters within *TBMetapopPy* change significance over time, with some of these corresponding to different stages of infection. Tables C.1, C.2 and C.3 show the effects of all parameters with a significant PRCC against each outcome, stratified into 10 days periods. We see that some parameters that are significant during the first 30 days (i.e. the primary stage of infection) lose significance at later stages. And conversely, parameters that are insignificant during the early stages

gain significance later in the simulated time. Examples of this include the rate of death of naïve T cells (μ_{TN}), where increased T cell death results in great bacterial numbers, but only during the primary infection, and enhanced recruitment of macrophages to the lymph nodes (β_{MY}), whereby increased cell recruitment there decreases the average size of lesions and the total number of bacteria within the model, but only from day 51 onwards.

6.5.3 eFAST

Whilst PRCC is an effective method of analysing model and parameter sensitivity for monotonic relationships between input parameters and model outputs, it is of limited use for cases where a non-monotonic relationship exists. In these cases of non-monotonic parameter/output combinations, non-significant PRCC values may be obtained despite the input parameter having influence on the model output [150]. Complex models with many parameters and many interacting compartment, such as *TBMetapopPy*, have a higher probability of introducing such non-monotonic relationships, and therefore it is advisable to perform further sensitivity analyses with methods such as eFAST which can handle these issues.

Whilst the eFAST method has significant utility, this comes at a cost: the number of simulations required in order to generate the necessary data is far higher than is needed for PRCC. The PRCC methodology is able to generate results for all parameters from every simulation, eFAST requires each parameter to be analysed in turn. The minimum number of runs for each parameter is 65 [208], and the resampling method requires this to be multiplied by N_R (the resample number). Furthermore, repeated simulations are needed to alleviate the aleatory uncertainty present within a stochastic model. The total number of simulations required (C) is shown in equation 6.9, where Rep is the number of repetitions of the parameter sample.

$$C = k * N_S * N_R * Rep \tag{6.9}$$

The computational cost of running eFAST was deemed to be too high for *TBMetapopPy*. With $k = 40$ (number of parameters) and taking $N_S = 65$, $N_R = 5$ and $Rep = 5$, this would result in 65,000 simulations - more than 32 times as many simulations as were performed for the PRCC analysis.

6.6 Conclusion

In this chapter, we have explored the role of UA and SA in model development, showing that as models become more and more complex, the risk of uncertainty from a variety of sources increases. In order to draw any significant conclusions from a model, it is of utmost importance that a modeller understands how uncertainties within the inputs of the model impact on the outputs of the model. We have described in detail two methods of performing SA on mechanistic models and have presented our framework designed to perform these on a model using parallel computation in order the time cost associated with these intensive analyses. We have verified our framework by comparing the results generated by it using an existing, deterministic model against previously published results for the same model, showing that we can achieve similar results. The benefit of our framework is the compatibility with an existing computational experiment control library, *epyc*, allowing for the separation of model definition and experiment procedure, meaning any model generated by a user have sensitivity analyses conducted on a remote, multi-core machine. This is particular important for SA tests, as fully exploring the possible parameter space involves a great many simulations, and running these in serial comes with a significantly large time-cost.

We have used our framework to perform sensitivity analyses on our *TBMetapopPy* model, the understand how the uncertainty present in the total number of bacteria, the number of lesions and the average lesion size can be attributed to uncertainty in the various parameters of the model. This is an important process for analysing our model: there exists a large amount of uncertainty within the parameters of our model, either

due to known heterogeneity in the natural systems (e.g. the large variation in macrophage numbers seen in different people [218]) or due to an incomplete mechanistic knowledge knowledge of the system (e.g. there are no published values concerning the rate of contact between individual cells with the body during a TB infection). Before trying to overcome this uncertainty, possibly through additional laboratory experiments or clinical trials focusing on these parameters, it is important to know which of the parameters affects the model outputs most. If a large amount of variation in the output can be attributed to an uncertain parameter, then this parameter should be prioritised in terms of further research which would aim the uncertainty in its value range.

Our results show that environmental attributes do contribute to the bacterial proliferation during an infection. The differential in perfusion from the apex to the base in particular has a large effect on both the total bacteria numbers and the average lesion size. The ventilation differential was not shown to have a significant effect on bacterial numbers, lesion numbers or lesion size during latency. This suggests that the apical localisation is less about oxygen availability and more about immune cell availability: the reduced perfusion at the apex of the lungs means the immune presence is weaker there, and thus presents bacteria with a more sheltered location in which to proliferate.

Furthermore, our results show the set of parameters with significant PRCC values differs during different stages of infection, lending credence to the notion that the immunological events that occur during primary and latency stages of infection are very different [67].

6.6.1 Limitations

The results presented in this chapter provide insight into the factors that are influential to the progression and scale of an infection within *TBMetapopPy*, but they are limited. As discussed, PRCC is incapable of

determining the sensitivity of an output due to a specific input in the case where a non-monotonic relationship exists between the variables. Within the context of a TB infection, it is reasonable to assume such non-monotonic relationships exist. Consider the case of inflammation: the natural response of the immune system to infection is to cause inflammation, usually to restrict the invading pathogen from accessing further areas of tissue. TB is no exception to this, with inflammation being an important part of granuloma control [77, 176]. However, inflammation is only useful to the host when carefully balanced [224]: excessive inflammation can lead to tissue destruction, and in the case of TB, might cause cavitation and thus afford bacteria access to highly-oxygenated airways [65]. Thus, the existence of non-monotonic relationships between host immune responses and bacterial outcomes necessitates the need of SA techniques which can handle such scenarios. However, as described in section 6.5.3, eFAST, a statistical test designed to handle non-monotonicities, comes at an exponentially greater computational cost. Whilst appropriate for smaller models of limited parameter numbers, running a complete eFAST analysis of a highly-parametrised model such as *TBMetapopPy* comes at too high a cost in terms of computation time. One solution would be to run simulations on multi-core machines or utilise cloud computing, but these are not always available to a modeller.

It is worth noting that what we present in this chapter is the determination of sensitivities within the model: we do not make the claim that these are definitively the contributing factors to infection within the real-world. Rather, they highlight possible avenues of research to explore in a laboratory setting; by performing SA to understand which parameters our model is most sensitive to, these parameters can then form the basis of *in vitro* and possible *in vivo* experiments which will help refine the parameter ranges, thus reducing the uncertainty of the model. This demonstrates the iterative nature of *in silico* model development and the need for collaboration between modellers and experimental scientists.

Chapter Seven

Conclusion

In which we discuss the findings of this thesis, detail the contributions, outline the limitations of our work and suggest possible future work based on this thesis.

7.1 Discussion

In this thesis, we have studied an emerging field within the wider of TB research: computational mechanistic modelling. Current efforts to reduce TB treatment duration (and thus improve adherence and reduce the threat of antibiotic resistance) have been unsuccessful, hampered by a lack of understanding of the complete set of complex dynamics that occurs during infection, even in spite of the disease having been present in society since antiquity. This lack of understanding is compounded by the lack of any animal model that completely mimics the pathophysiology seen in humans. Computational models offer a solution to this issue: by using a synthetic human lung, we can replicate the dynamics as seen in human disease, with recent advancement in computing power and techniques now allowing us to build more complex models.

In chapter 2, we reviewed the literature on TB infection within humans, looking at the series of complex interactions between host and pathogen that lead to a disease with varying stages: from primary infection arising after initial bacterial deposition, to the establishment of a latent form of disease whereby the bacteria are contained but are typically not eradic-

ated, with the host remaining in an asymptomatic and non-contagious state for years or decades, through to the degradation of containment leading to the emergence of a post-primary form of disease. In particular, we noted that the localisations within the lung differ these different stages, with primary infection typically starting in the lower regions of the lung, and post-primary disease occurring in the apices of the lung. As post-primary disease is the stage where the vast amount of disease transmission occurs, and which makes up the majority of clinical cases, understanding how and why bacteria move to the apices during a TB infection is of paramount importance. In particular, creating novel treatments that will be effective against the bacteria requires an understanding of the environments in which the bacteria reside. It is believed that environmental heterogeneities such as differentials in oxygen tension and blood perfusion contribute to creating a micro-environment within the apex of a lung that benefits bacterial proliferation, but it remains unclear as to what extent each of these environmental heterogeneities impacts the disease.

Previously, *in silico* mechanistic models of TB have focused on modelling the disease at a microscopic level, typically to investigate the factors that contribute to the formation, stabilisation and/or degradation of a single granuloma. This work plays a vital part in improving the understanding of TB, but comes with a limitation: modelling a TB life-cycle whereby the breakdown of the initial lesion is the contributing factor of post-primary disease ignores the evidence that the entry lesion and exit lesion are in two distinct locations in the lung, implying that some form of bacterial dissemination must occur, and that the heterogeneity of conditions within the lung induces bacterial growth at the apex. Prior to this thesis, very few existing computational models of the disease had examined TB as it occurs over the whole lung, and of those that had, none had included these environmental heterogeneity phenomena.

On this basis, we framed our research questions as:

- To what extent can we use computational models of whole organ

scale to provide insight into the dynamics that occur during TB infection within the lungs?

- What environmental factors within the lung contribute to the specific localisations of TB during different stages of the disease?

7.2 Contributions

In order to answer the research questions above, this thesis makes a number of contributions, which we discuss in greater detail here:

7.2.1 A framework for the creation of networked metapopulation models which incorporate environmental heterogeneity

The main premise of our research was to build the first computational model of TB to incorporate environmental heterogeneity in order to study its impact on the disease pathophysiology. In chapter 3, we laid the groundwork for this model by first introducing a modelling framework, *MetapopPy*, that allows a modeller to create a networked metapopulation model that incorporates environmental heterogeneity. This is achieved by taking a network and attaching environmental attributes to each individual node. Then, a series of events can be programmed to occur, some of which may be dependent on the environmental attributes, thus creating a heterogeneous system of dynamics whereby the rates of occurrence of events may differ depending on where in the network they are occurring. This framework removes any responsibility of change propagation from the user: when a change is made that affects the environment or population at each patch of the metapopulation, the changes are automatically propagated through to the rate calculations, only requiring the user to define how event rates should be calculated from a patch. This framework is abstract in its nature and can be applied to a wide variety of fields; we demonstrate through our validation, creating epidemic models of disease using the framework and comparing the results retrieved from them against previously published results.

7.2.2 A whole organ model of a pulmonary TB infection that incorporates the environmental heterogeneity found within the lung to determine the role it plays upon disease progression

In chapter 4, we then outlined how this framework was utilised to create *TBMetapopPy*, the first TB computational model including environmental heterogeneity within the lung. A network of patches, representing the bronchopulmonary segment of the lung and the associated lymph nodes, was created and the lung patches were assigned environmental variables corresponding to known differentials in the lung that are believed to affect to TB. Nodes where connected to indicate the possibilities of bacterial and immune cell dissemination. On top of this network, a set of events simulating the various translocations and interactions of both and immune cells between each other and the environment, allowing the differentials in the environment to impact on factors such as immune cell recruitment and bacterial state change.

Our results showed that the introduction of heterogeneity had profound impacts on the bacterial loads present during infection, whereby creating regions within the network of reduced perfusion and increased oxygen allowed for overall bacterial proliferation at a rate far higher than in a similar model with a homogeneous environment. In doing so, we showed that in order to fully appreciate the scale of disease within a host during TB infection, environmental heterogeneity must be accounted for in simulation models.

Building on this initial model, we developed a secondary model with a more granular network, as noted in chapter 5. We devised a space-filling tree algorithm that creates a branching structure similar to that seen in the human bronchial tree. This algorithm then formed the basis of our new network, where each end point of the tree corresponded to a single division of the total alveolar tissue within the lung. This granular network allowed us to more accurately reflect the differentials of ventilation

and perfusion as linear scales (and subsequently oxygen tension as an exponential scale) running across the height of the lung, and we therefore applied environmental attributes based on the vertical position of each patch. The events present and the parameters used for the model were then adjusted to this paradigm.

Our results showed further proof of how environmental heterogeneity can impact disease progression: with the initial infection in the lower regions spreading to patches across the lung environment, but at latency establishing different bacterial loads in different locations. Our findings show an ideal region of the lung exists, approximately one quarter of the height of the whole lung from the apex, whereby perfusion is high enough to permit bacterial dissemination but the environmental conditions lead to establishment of larger lesions than in any other region. We therefore establish a heterogeneity of lesion size, and show that have shown that a drop in t-cell recruitment will lead to the lesions at the apex reaching similar levels to those seen during the primary stage, effectively establishing a post-primary form of disease.

7.2.3 A software framework to efficiently perform sensitivity and uncertainty analysis

In chapter 6, we outline the importance of uncertainty and sensitivity analysis in understanding how uncertainty within the inputs of the model propagates to create uncertainty in the output, and outline some existing statistical techniques that have previously been used in published articles to apportion the output uncertainty to the various uncertain inputs of a system. This is of particular importance for TB models, where the lack of a complete mechanistic knowledge and the unavailability of relevant data to some of the dynamics of infection can lead to large numbers of uncertain parameters. Only by understanding how this uncertainty affects the model can we draw any valid conclusions. Performing these sensitivity analyses requires a wide sample of the parameter space to be explored, sometimes using a complex algorithm, and thus a large number of simulations need to be run. Performing these in serial can be

at best time-consuming and at worst completely unfeasible. Therefore, we developed a framework, using an existing experiment control library, to automatically create the parameter samples of two established sensitivity tests and to distribute these samples across a parallel compute server to improve performance. We again verified our code against previously published results.

7.2.4 Insight into biological factors that contribute to TB progression

We then used this framework to perform sensitivity analyses on our *TBMetapopPy* model, in order to determine which of our parameters and dynamics were contributing to the variety of disease outcomes that were shown in our results (in terms of lesion size and total bacteria).

Whereas many sources have attributed the apical localisation of TB solely to the increased oxygen tension at the apices, we have shown that of all the environmental heterogeneities within our model, it is perfusion that has the greatest impact on the lesion size. This affects both the oxygen availability and the immune response throughout the lung.

T cell functions have long been known to be critical in controlling TB infection [47, 78], and our sensitivity results corroborate this, showing the importance of T cells in containing infection, with most T cell-related parameters have a significant PRCC value against our outputs.

Understanding where bacteria reside during infection is critical to the creation of new treatment regimens, as different drugs have different efficacies in different locations. Our sensitivity results show that during latency, it is extracellular replication that has more of an impact on our disease results than intracellular replication, indicating that the killing of extracellular bacteria may be of more importance than killing those that are intracellular.

It is already known that bacteria disseminate across the lung and lymph-

atics during infection, but the mechanisms by which they do this are unclear. Macrophage migration to the lymphatics is rarely considered as a contributing factor, due to the low rate at which it occurs, but our sensitivity results indicate that this is a significant factor in disease outcomes, far more so than the migration of dendritic cells. But our results suggest that the transfer of macrophages can improve disease outcome, leading to smaller lesions and less total bacteria, possibly due to the draining of bacteria away from areas of pathology in the lung and into the lymphatics where a stronger T cell response can be formed. Our research suggests that this mechanism should be explored in greater detail, perhaps in a laboratory setting.

Our findings show not just the factors which affect disease outcomes, but also the times during infection at which they are significant. If we are able to determine the complete set of significant parameters during each stage of infection, we can develop a regimen which specifically targets these parameters at each stage, focusing on the areas that accelerate bacteria growth at each point to better reduce the chance of a negative outcome and thus improve the speed at which bacteria are destroyed.

7.3 Limitations

Systems biology approaches, of the kind utilised throughout this thesis, aim to replicate the real world and as such they are built upon our current understanding of the systems, which raises an important question:

> Will our models not be inherently flawed from the outset if our current knowledge is wrong?

Indeed, all models are 'wrong': they are abstractions of the real world designed to help us make sense of it. But this does not preclude them from being useful, so long as we understand why they are 'wrong'. This requires verification and validation: verification ensures that the model that has been built matches our design specifications (i.e. our current understand of the disease dynamics) correctly, whilst validation of a model checks that the model produces results that reflect the real world

[223]. If our model is well-verified, but we are unable to validate our model predictions against real-world data, then this suggests that our current understanding is incorrect. The model can be refined, moving it away from the currently accepted paradigm and towards one that better reflects available evidence. The discrepancies between the new model and the current understanding can then provide the basis for *in vitro* and *in vivo* investigation. In this sense, modelling allows us to gain insight into disease and treatment dynamics and provides evidence in support (or in opposition) of our beliefs about disease mechanics. Close collaboration with experimentalists and trialists is needed for this, as all model outcomes must be viewed through the lens of clinical disease to test the plausibility of predictions.

The relative paucity of knowledge regarding the processes of a disease such as TB presents a challenge when developing mechanistic models:

> If we don't know how the disease develops in the body, how do we determine realistic parameter values for the model?

All parameters used within a model will be based on biological or clinical experiments that are inherently variable and can have wide confidence intervals - it is challenging to understand the impact of different experimental definitions and methods on the results produced. This can make reliable parameter values hard to assign and further highlights the importance of working closely with experimentalist collaborators. In our model, a wide number of parameters had no existing values we could derive from the literature, requiring us to estimate possible values and thus introducing greater uncertainty. One way to mitigate this difficulty is through sensitivity analysis, as we have demonstrated in Chapter 6.

However, these analyses can be computational intensive, even when run in parallel, as we showed with regards to running eFAST analyses to account for non-monotonic relationships. This represent the classic trade-off encountered when modelling: the model must be complex enough to represent as much of the dynamics of the real-world system as needed for the problem being investigated, but not so complex as

to become impossible to draw any meaningful conclusions from, either due to extensive computational costs when running simulations or the introduction of complex interactions between parameters. Our model is relatively complex and yet does not incorporate many factors that may play a large role in bacterial dissemination, including local transmission along the bronchial tree. These events are not particularly relevant to our investigations, but may need to be included in future work.

Compounding the issue of choosing an appropriate level of model complexity is the fact that, currently, the *MetapopPy* framework is not efficient when run on networks with large numbers of nodes and/or events due to the large number of rate calculations needed, limiting the number of repetitions permissible and thus retaining aleatory uncertainty. Therefore, future iterations of the framework should look to improve the data structures used to store rates and the mechanisms by which changes are propagated.

The aim of this thesis is not the development of a model that can be used to predict patient outcomes from novel treatments. Doing so would require a complete knowledge of the underlying dynamics and, as is discussed throughout this thesis (particularly in Chapter 2), our current understanding of TB is not nearly comprehensive enough to permit this. Instead, the intention of this thesis is to provide insight: to improve our current understanding by simulating those events that we are uncertain about, either providing an explanation for currently unexplained dynamics, or to highlight the direction that further research (either computational or in a laboratory or clinical trial) should progress in.

7.4 Future work

The results presented in this thesis are interesting in their own right, but are necessarily incomplete due to the limitations listed in this chapter and in the previous chapters. In this section, we extrapolate possible future avenues of research which build upon the work presented here.

7.4.1 Validation of results

We have shown that environmental heterogeneities within the lung cause an apical localisation of lesions that occur slightly below the extreme apex. Whilst this agrees with the general consensus of the localisation of post-primary TB [11, 67, 94, 173], further validation should be performed to match the results shown here to patient x-rays for untreated TB, to determine the accuracy of predictions in terms of the exact distance of the major lesions from the very apex of the lung.

7.4.2 Enhancements to the current *TBMetapopPy*

Here, we list possible factors that could be incorporated into new iterations of the *TBMetapopPy* model and further experiments that can be run using the model.

7.4.2.1 Lung anatomy and morphology

The current network topology used to model the lungs is simplistic: we have represented the lung as a series of two-dimensional points built upon a space filling-tree that fills a rectangle. The reality is that the lungs have a non-uniform shape, with the apex of the lung being much more rounded and narrower than the wide base of the lung. This could have significant impact on the pathophysiology of TB: if the apices also have less space available than the basal regions this could present a greater challenge to *M. tuberculosis* bacteria in finding a preferential niche. Shape is also significantly different between left and right lungs, with the left lung being smaller to accommodate the heart in that side of the chest cavity. Incorporating a second lung, with a different topology, and performing analyses to understand how the differences between lungs impacts disease will be an important avenue of research.

TB disease obviously occurs in three-dimensions. Our current model only accounts for two-dimensional environmental changes across the lung as these are the most common concepts that affect disease in the TB literature. A three-dimensional model could show interesting dynamics,

particularly in instances where the lung orientation does not remain constant (see section 7.4.2.2).

In our model, we have made the assumption that all disease that occurs within the lung does so at the alveolar tissue. Whilst a useful abstraction, this may be simplistic, particularly when trying to understand the means by which bacteria disseminate across the organ. Endobronchial TB (TB disease of the bronchial tree) is known to occur [215], and it has been hypothesised that bronchial obstruction can be a crucial factor in the pathogenesis of cavities in TB [120].

7.4.2.2 Dynamic lung environment

Our lung model contains a single, static lung, but the reality is that no lung is ever static. Various different factors influence the shape of the lung at a given time and the environmental heterogeneities that occur there.

The environmental heterogeneities within our lung model are based on those exhibited in the human lung due to gravity when the lung is upright (i.e. the person is stood/sat straight up). But changes in position can affect the scale and/or direction of differentials in ventilation, perfusion and thus also oxygen tension: when lying down in a supine position, more blood is sent to the entire posterior region of the lung, thus increasing the blood at (part of) the apices. This may explain the evidence that bed rest (which would necessarily involve the patient spending more time lying down) had previously been prescribed as a cure for TB [173]. Exercise and changes in altitude also alter the blood distribution within the lungs, increasing the blood flow to the apices in order to improve oxygen exchange when it is most needed [234, 235]. As human beings are never constantly static (particularly not for the long periods of time a TB infection takes place), models that display dynamic changes in their environmental heterogeneities would better reflect the real-world, and could test the effects of bed rest when combined with pharmacological treatment.

7.4.2.3 Between-person heterogeneity

Lung morphology and physiology are affected by a number of factors, including height and body size [48], and sex [18, 45, 222]. Differences by sex may be of particular interest to TB researchers, as there has been shown to be gender differences in responses to treatment: an analysis of the REMoxTB clinical trial (see section 2.10.3) showed men with cavitation responded poorer to treatment than women with or without cavitation [172]. These differentials in disease outcome between genders may be caused by biological sex differences between male and female lung morphologies [45, 179], although the causes may also be socio-economic or cultural [118]. Further experiments using *TBMetapopPy* should focus on investigating how differentials between people influence clinical outcomes in order to determine the patients who will be at greater risk of relapse.

7.4.2.4 Immune cell enhancements

The interactions that occur between immune cells and bacteria during a TB infection are many and various, and in order to keep *TBMetapopPy* parsimonious we have necessarily had to make restrictions on the number and type of immune cells we include, but we acknowledge that there are more complexities to the human immune system that should also be included in any future iterations of the model.

In *TBMetapopPy*, we include just two states for T cells: naïve and activated. In fact, T cells populations contain a variety of phenotypes. One major distinction that is not included in the current iteration of the model is the difference between CD4 (effector) T cells and CD8 (cytotoxic) T cells. In our model, all activated T cells perform actions of both of these phenotypes, but in order to get a better understanding of how environmental differentials impact distinct sub-populations of cells, future iterations of the model could incorporate this and other distinctions within the existing model compartments.

The phagocytes we have included with *TBMetapopPy* are limited to

macrophages and dendritic cells, based on previous*in silico* models of within-host TB dynamics. However, as previously stated, it is known that neutrophils play a major role in the disease pathology as they are often the first cells to encounter bacteria [145]. Whilst the roles that neutrophils play during infection are largely covered by the current phagocytes within the model, separating neutrophils out into a separate compartment may allow for a more accurate representation of the real world dynamics, and to better understand cell-to-cell interactions and the impact the environment has on these, as neutrophils have strong interactions with macrophages [145].

Whilst classically the role of humoral immunity in TB was thought to be limited [40], B cells and antibodies have been re-appraised to better understand how they might influence disease outcomes [1, 146]. *TBMetapopPy* could be a useful tool to model the humoral and cell mediated immune systems and understand their interactions and how they might be targeted to improve patient response to disease and treatment.

7.4.2.5 Bacterial heterogeneities

TB exists in a variety of different strains, with some patients being co-infected with more than one strain [19, 230]. It is reasonable to assume these strains may behave differently (different replications rates, differential probability of destruction by immune cells, differential resistance to antibiotics, etc.), and thus *TBMetapopPy* could be a useful tool to understand how differences in the mycobacteria affect the pathology over the whole organ. Another bacterial heterogeneity that could be incorporated into different compartments of bacteria is that of drug resistance, with different compartments having differing levels of resistance to certain drugs (see 7.4.3 for further discussion on the causes and implications of a model with drug resistance).

7.4.2.6 Impacts of the disease on the lung

Our current iteration of *TBMetapopPy* has focused on how the environment impacts the disease, but has not yet investigated the reverse: how

TB pathology can alter the lung environment. This can occur during infection in many ways, from obstruction of the bronchi caused by a pneumonia creating an abundance of foamy macrophages and thus a dysregulation of the immune response in the region [120], to cavitation, where large destruction of the bronchial tissue affords bacteria access to regions of greater oxygenation. In the latter event, this may require a restructuring of the *TBMetapopPy* network, where cavitation causes new patches (which are close-by in physical space but perhaps a long distance via the bronchial tree) to be connected.

7.4.3 Pharmacology modelling

Understanding how treatment of TB is affected by the environment requires us to understand how the disease exists differently in different regions within the lung. As such, our model has been confined to studying just an untreated version of TB, but having made such a model, a logical next step would be to introduce a treatment component to understand how the heterogeneities of the environment impact the pharmacology of TB, either by the environment creating different populations and conditions for bacteria which cause heterogeneities of treatment, or through the environment directly impacting the way chemotherapy is delivered to the areas of infection.

All chemotherapy for TB enters the lung through the blood (after having been digested), and therefore it is reasonable to assume that the delivery of the drugs to the alveolar tissue is impacted by heterogeneities in the lung, in particular by the differential in blood perfusion. We have shown how the apex of the lung is preferential in an untreated case of TB, by reduced access of immune cells, but the same may also be true for treatment, with reduced perfusion resulting in levels of drugs that do not reach the MIC levels necessary to suppress and/or eliminate bacteria. Therefore, a future iteration of our model could investigate if environmental heterogeneities in the lung can lead to inadequate treatment and possibly even allow for prediction of drug levels that would be sufficient. It is also possible that differences in the environment may impact of the

PD of the drugs, with different drugs having different activity depending the environment which they reach (i.e. the differentials of oxygen and ventilation, or other heterogeneities know to exist in the lung such as pH, may alter the effectiveness of each drug), and this could investigated by our model.

7.4.3.1 Emergence of drug resistance

Exposure to insufficient quantities of drugs can lead to development of drug resistance in the remaining bacteria that are not destroyed, and this presents a global danger through the emergence of MDR-TB and XDR-TB. This could be incorporated into our model, by including drug-susceptible and drug-resistant compartments of bacteria, and allowing movement between these based on the exposure levels of bacteria.

7.4.4 Virtual clinical trials

By extending our model to include:

- the ability to create a heterogeneous population of varying lung environments,
- pharmacological components, including drug activity and differential treatment regimens and adherences,

then these components combined effectively create a 'virtual clinical trial', whereby a synthetic population of different demographics (representative of the demographics found in real life) could be created, and possible new treatment regimens could be created and trialled over this population. If achieved, this would be a major benefit to the TB research community: the ability to predict how likely a novel regimen would be to succeed when put to clinical trial would allow for better prioritisation of many possible new regimens, reducing the risk of clinical trial failure and saving both time and money and the hastening our search for an improved, shorter treatment. Whilst this may be a long way off, we believe the work presented in this thesis is an important step on the road towards this goal of a virtual clinical trial.

7.4.5 Enhancements to the current sensitivity analysis framework

The scope of the sensitivity analysis framework we presented here was limited by time and the requirements of the sensitivity analysis required for the *TBMetapopPy* model. There are many more advancements that could be made to this framework to enable users to quickly and conveniently perform their own analyses, without requiring a comprehensive understanding of the mathematical processes involved in creating the samples and analysing the results[1]. The first of these would be to incorporate the framework directly into the `epyc` library: `epyc` currently only allows for the creation and running of experiments. Incorporating our sensitivity analyses into this would make the entire analysis process seamless, requiring only one library to run everything. Secondly, further methods of sensitivity analysis should also be included, allowing the user to run different analyses dependent of the type of simulation they are investigating.

[1] This is most evident in eFAST - the processes required to derive the samples (sampling each parameter at various frequencies including resamples and random phase shifts), would be difficult for researchers without a mathematical background to comprehend. Making these processes simpler and only requiring the user to specify parameters and their ranges would be a significant benefit.

Appendix A

Secondary model parameters

Parameter	Description	Value	Ref
λ_R	B_{ER} replication rate	0.814	[26]
λ_D	B_{ED} replication rate	0.26	[26]
λ_I	B_{IM} replication rate	0.26	E
σ_λ	Sigmoid for B_{IM} replication	2.0	[153]
ϕ	Carrying capacity of macrophages	50	[153]
ξ	Rate of conversion between bacterial states	1	E
θ_ξ	Half-saturation for conversion between bacterial states	1	E
σ_ξ	Sigmoid for conversion between bacterial states	2	E
τ_{BY}	Rate of translocation of bacteria from the lymphatics to the lung	1e-3	E
$\theta_{\tau BY}$	Half-sat of caseum preventing translocation of bacteria from the lymphatics into the lung	500	E

Table A.1: Bacterial event parameters within *TBMetapopPy* (E = estimated)

Parameter	Description	Value	Ref
α_D	Rate of standard recruitment of D_I into the lung	599e4	E
β_D	Rate of enhanced recruitment of D_I into the lung	5000.05	E
$\theta_{\beta D}$	Half-saturation value for enhanced recruitment of D_I into the lung	5500	E
μ_{DI}	Death rate of D_I	0.01	[153]
μ_{DM}	Death rate of D_M	0.3	[153]
γ_D	Rate at which D_I encounters bacteria	0.3	E
$\theta_{\gamma D}$	Half-saturation value for contact between D_I and bacteria	5500	E
τ_D	Rate of translocation of D_M from lung to lymphatics	0.55	[154, 202]

Table A.2: Dendritic cell event parameters within *TBMetapopPy* (E = estimated)

Parameter	Description	Value	Ref
α_{ML}	Rate of standard recruitment of M_R into the lung	599e5	[218]
β_{ML}	Rate of enhanced recruitment of M_R into the lung	50000.05	E
$\theta_{\beta ML}$	Half-saturation value for enhanced recruitment of M_R into the lung	5e3	E
α_{MY}	Rate of standard recruitment of M_R into the lymphatics	53.465	[154]
β_{MY}	Rate of enhanced recruitment of M_R into the lymph	750	E
$\theta_{\beta MY}$	Half-saturation value for enhanced recruitment of M_R into the lymph	5500	E
ω	Weighting value for chemokine release by M_I to M_A	0.55	E
μ_{MR}	Death rate of M_R	0.01	[153]
μ_{MA}	Death rate of M_A	0.01	[153]
μ_{MI}	Death rate of M_I	0.01	[153]
ψ_{μ}	Percentage of bacteria destroyed during natural death of M_I	0	E
ζ	Rate of bursting of M_I	0.275	[153]
ψ_{ζ}	Percentage of bacteria destroyed during bursting of M_I	0	E
τ_M	Rate of translocation of M_I from the lung to the lymphatics	0.001	E

Table A.3: Macrophage event parameters within *TBMetapopPy* (E = estimated)

Parameter	Description	Value	Ref
ϵ_{MB}	Rate of activation of M_R by extracellular bacteria	0.01	E
$\theta_{\epsilon MB}$	Half-saturation for activation of M_R by extracellular bacteria	5500	E
ϵ_{MT}	Rate of activation of M_R by T_A	0.3	[153]
$\theta_{\epsilon MT}$	Half-saturation for activation of M_R by T_A	5500	E
κ_M	Rate at which T_A destroy M_I	1.35	[153]
$\theta_{\kappa M}$	Half-saturation value for T_A destruction of M_I	1e3	E
$\psi_{\kappa M}$	Percentage of bacteria destroyed when M_I is destroyed by T_A	0.5	E
γ_R	Rate at which M_R ingest bacteria	0.3	[153]
$\theta_{\gamma R}$	Half-sat for contact between M_R and bacteria	5500	E
η	Probability M_R becomes infected when ingesting bacteria	0.75	E
γ_A	Rate at which M_A ingest bacteria	0.8	E
$\theta_{\gamma A}$	Half-sat for contact between M_A and bacteria	5500	E

Table A.4: Parameters for events involving macrophages coming into contact with other cells within *TBMetapopPy* (E = estimated)

Parameter	Description	Value	Ref
α_T	Rate of standard recruitment of T_N into the lymphatics	1000	[153]
β_T	Rate of enhanced recruitment of T_N into the lymphatics	1000	E
$\theta_{\beta T}$	Half-saturation value for enhanced recruitment of T_N into the lymphatics	1e3	E
ϵ_T	Rate of activation of t-cells	0.4	[153]
$\theta_{\epsilon T}$	Half-saturation value for activation of t-cells	1e3	E
τ_T	Rate of translocation of t-cells from lymphatics into the lung	0.625	[153]
σ_T	Sigmoid for translocation of t-cells from lymphatics into the lung	0.25	E
μ_{TN}	Rate of death for T_N	0.102	[153]
μ_{TA}	Rate of death for T_A	0.333	[153]
λ_T	Rate of replication for T_A	0.15	E

Table A.5: T-cell event parameters within *TBMetapopPy* (E = estimated)

Appendix B

Model parameters for sensitivity analysis

Parameter	Range	Distribution
S_V	2.0 (0.1)	N
S_Q	3.0 (0.1)	N
S_G	1.0-5.0	U

Table B.1: Environmental parameters within *TBMetapopPy* for sensitivity analysis. In Distribution, U = Uniform, N = Normal. For uniform, a range is given, for normal, mean and standard deviation are given.

Parameter	Range	Distribution
λ_R	0.814 (0.03)	N
λ_D	0.256 (0.01)	N
λ_I	0.256 (0.01)	N
σ_λ	2.0 (0.01)	N
ϕ	50 (10)	N
$\tilde{\xi}$	0.01 - 2.0	U
$\theta_{\tilde{\xi}}$	0.01 - 2.0	U
$\sigma_{\tilde{\xi}}$	1.0 - 3.0	U
τ_{BY}	5e-4 - 2e-3	U
$\theta_{\tau BY}$	250 - 750	U

Table B.2: Bacterial event parameters within *TBMetapopPy* for sensitivity analysis. In Distribution, U = Uniform, N = Normal. For uniform, a range is given, for normal, mean and standard deviation are given.

Parameter	Range	Distribution
α_D	599e4 (75e4)	N
β_D	1e-1 - 1e5	U
$\theta_{\beta D}$	1e3 - 1e4	U
μ_{DI}	0.01 (4e-3)	N
μ_{DM}	0.3 (0.1)	N
γ_D	0.2 - 0.4	U
$\theta_{\gamma D}$	1e3 - 1e4	U
τ_D	0.1 - 1.0	U

Table B.3: Dendritic cell event parameters within *TBMetapopPy* for sensitivity analysis. In Distribution, U = Uniform, N = Normal. For uniform, a range is given, for normal, mean and standard deviation are given.

Parameter	Range	Distribution
α_{ML}	599e5 (75e5)	N
α_{MY}	53.465 (3.0)	N
β_{ML}	1e-1 - 1e5	U
$\theta_{\beta ML}$	1 - 1e4	U
β_{MY}	600 - 900	U
$\theta_{\beta MY}$	1e3 - 1e4	U
ω	1e-2 - 1	U
ϵ_{MB}	1e-4 - 0.1	U
$\theta_{\epsilon MB}$	1e3 - 1e4	U
ϵ_{MT}	0.1 - 0.5	U
$\theta_{\epsilon MT}$	1e3 - 1e4	U
μ_{MR}	0.01 (4e-3)	N
μ_{MA}	0.01 (4e-3)	N
μ_{MI}	0.01 (4e-3)	N
ζ	0.25 (0.05)	N
κ_M	0.7 - 2	U
$\theta_{\kappa M}$	1e1 - 2e3	U
$\psi_{\kappa M}$	0.5 - 1	U
γ_R	0.3 (0.01)	N
$\theta_{\gamma R}$	1e3 - 1e4	U
η	0.5 - 1.0	U
γ_A	0.2 - 1.4	U
$\theta_{\gamma A}$	1e3 - 8e3	U
τ_M	1e-4 - 1e-1	U

Table B.4: Macrophage event parameters within *TBMetapopPy* for sensitivity analysis. In Distribution, U = Uniform, N = Normal. For uniform, a range is given, for normal, mean and standard deviation are given.

Parameter	Range	Distribution
α_T	1000 (3)	N
β_T	900 - 4000	U
$\theta_{\beta T}$	1 - 2e3	U
ϵ_T	0.1 - 0.7	U
$\theta_{\epsilon T}$	10 - 2e3	U
τ_T	0.3 - 0.95	U
σ_T	1e-3 - 3	U
λ_T	1e-3 - 0.3	U
μ_{TN}	0.102 (1e-2)	N
μ_{TA}	0.333 (1e-2)	N

Table B.5: T-cell event parameters within *TBMetapopPy* for sensitivity analysis. In Distribution, U = Uniform, N = Normal. For uniform, a range is given, for normal, mean and standard deviation are given.

Appendix C

PRCC results

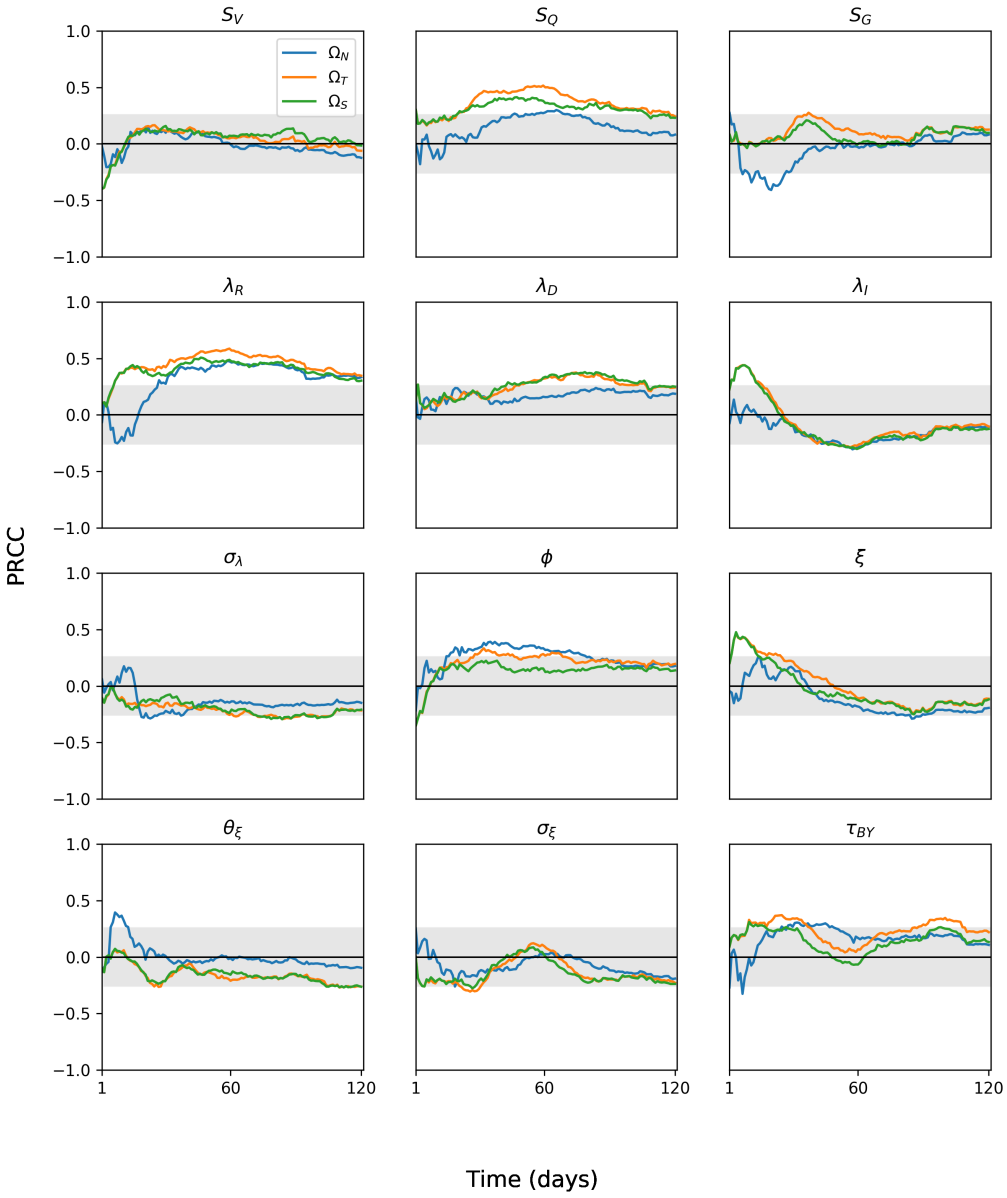


Figure C.1: Sensitivity plots for various parameters within *TBMetapopPy*, plotting the PRCC value of the parameter against the three outputs over time. Grey shaded area shows non-significance (p<0.01).

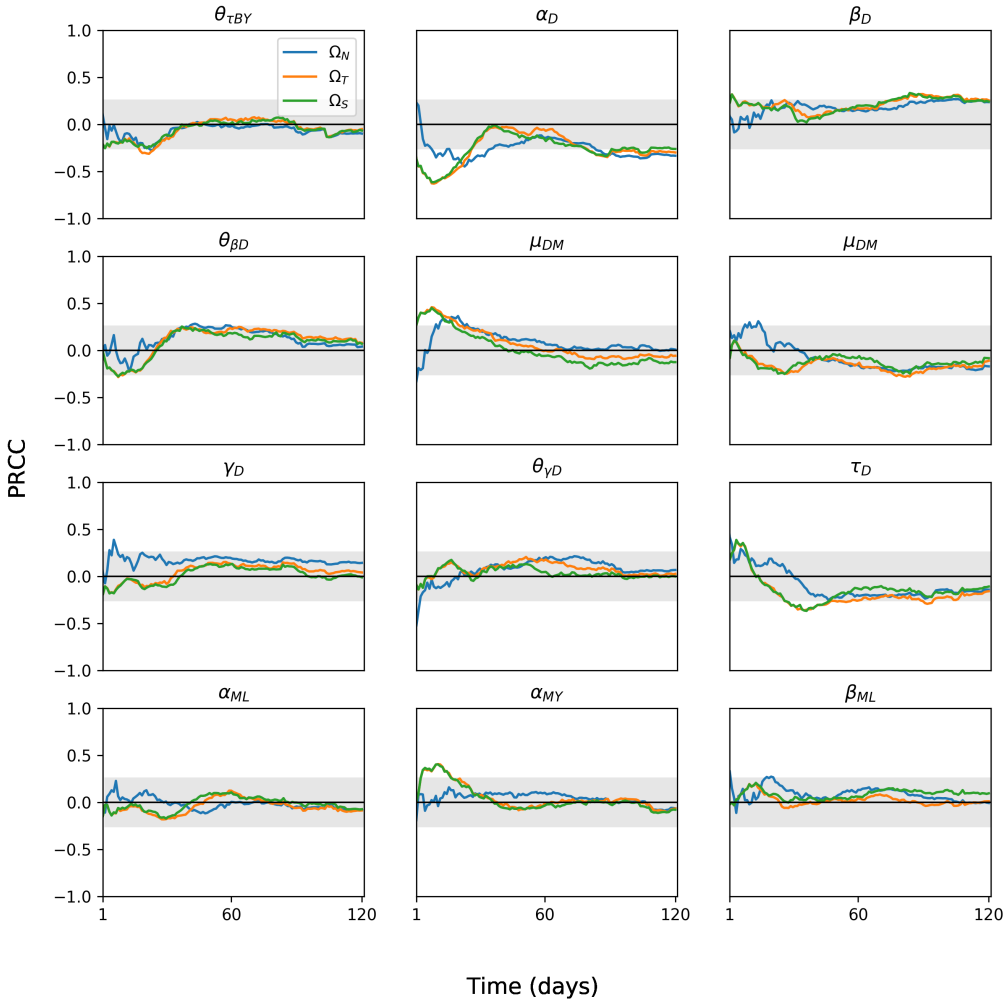


Figure C.2: Sensitivity plots for various parameters within *TBMetapopPy*, plotting the PRCC value of the parameter against the three outputs over time. Grey shaded area shows non-significance (p<0.01).

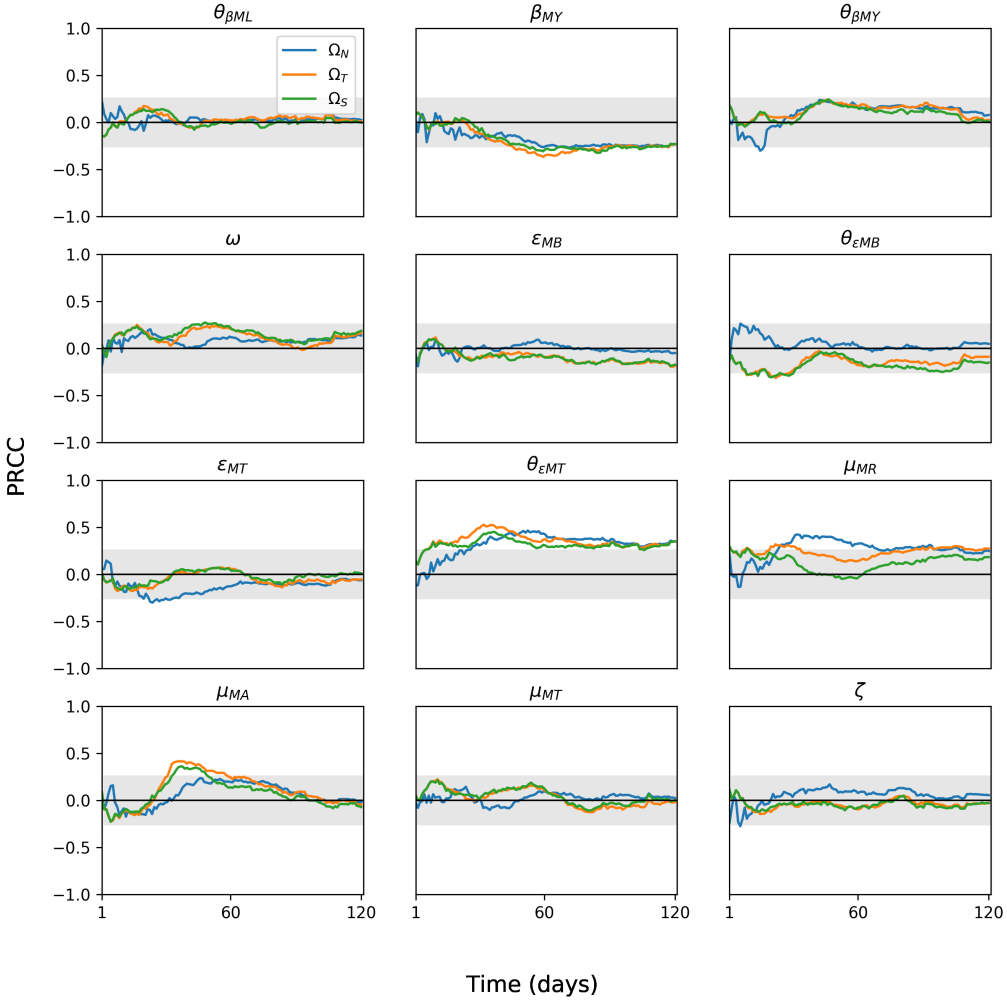


Figure C.3: Sensitivity plots for various parameters within *TBMetapopPy*, plotting the PRCC value of the parameter against the three outputs over time. Grey shaded area shows non-significance (p<0.01).

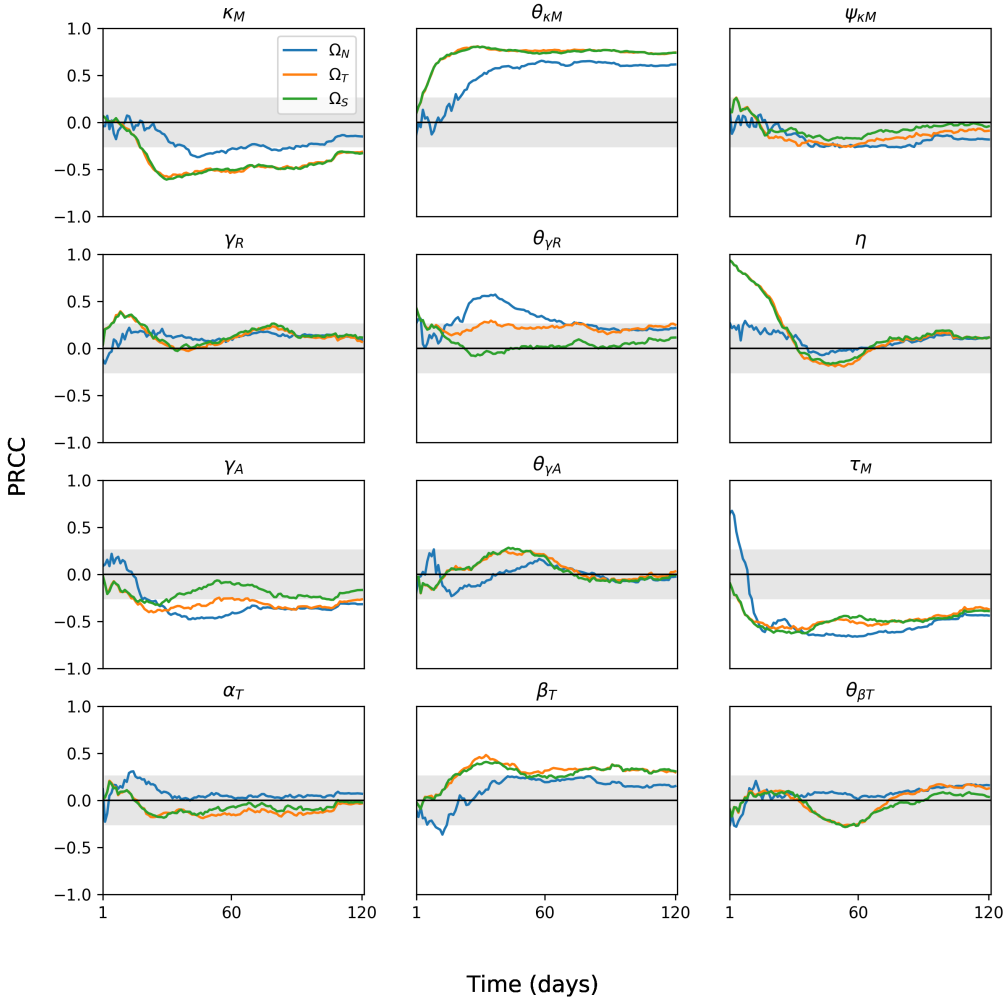


Figure C.4: Sensitivity plots for various parameters within *TBMetapopPy*, plotting the PRCC value of the parameter against the three outputs over time. Grey shaded area shows non-significance (p<0.01).

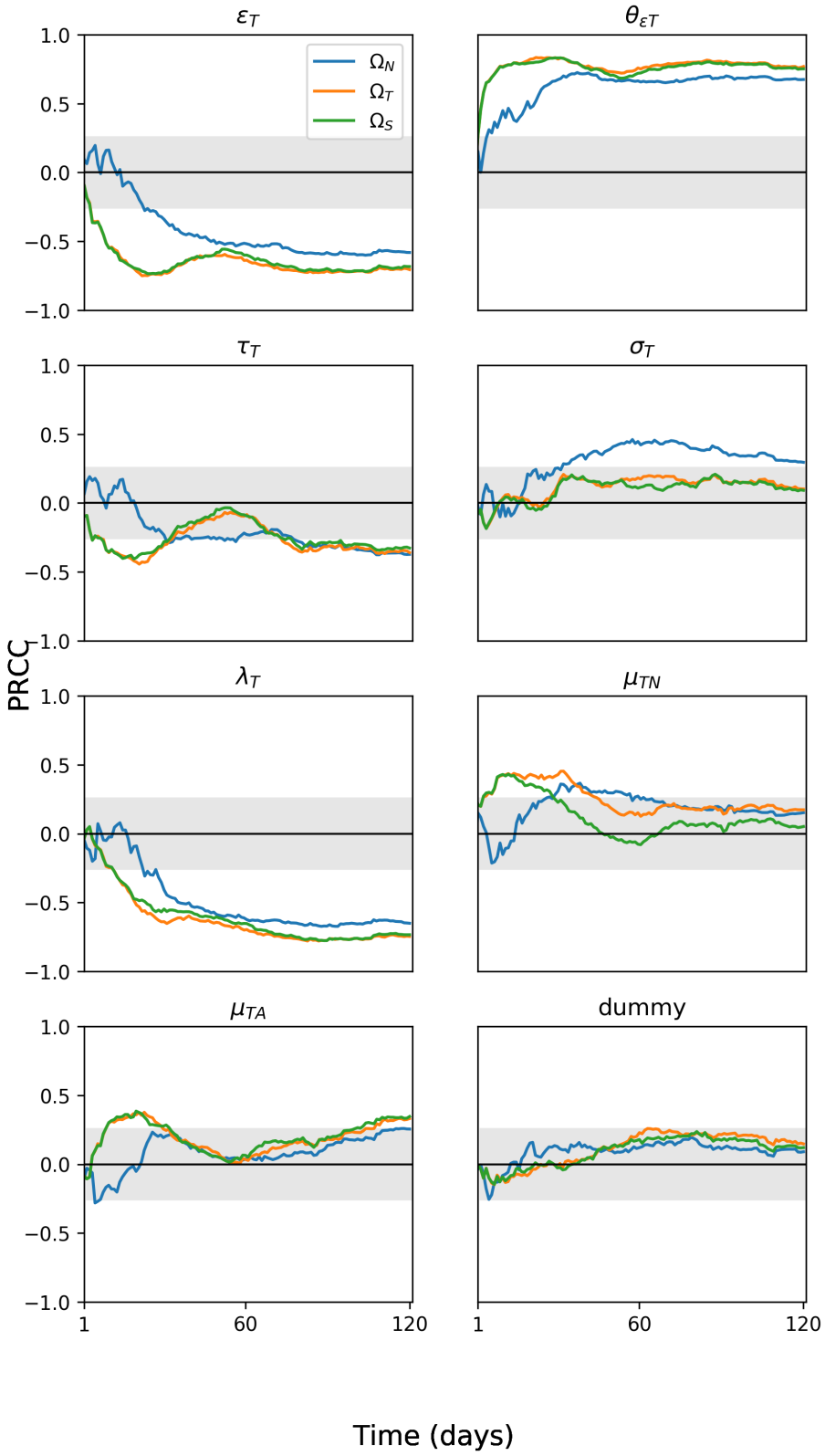


Figure C.5: Sensitivity plots for various parameters within *TBMetapopPy*, plotting the PRCC value of the parameter against the three outputs over time. Grey shaded area shows non-significance (p<0.01).

	Days after initial infection											
	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	111-120
S_V	-											
S_Q	+		+	+	+	+	+	+	+	+	+	+
λ_R	+	+	+	+	+	+	+	+	+	+	+	+
λ_D	+				+	+	+	+	+	+	+	+
λ_I	+	+			-	-	-					
σ_λ								-	-	-	-	
ϕ	-											
ξ	+	+										
θ_ξ											-	-
σ_ξ	-		-									
τ_{BY}		+	+							+		
α_D	-	-	-					-	-	-	-	-
β_D	+							+	+	+	+	
$\theta_{\beta D}$	-											
μ_{DI}	+	+										
τ_D	+		-	-	-							
α_{MY}	+	+										
β_{MY}						-	-	-	-	-	-	-
ω					+	+						
$\theta_{\epsilon MB}$		-	-	-								
$\theta_{\epsilon MT}$	+	+	+	+	+	+	+	+	+	+	+	+
μ_{MR}	+	+										
μ_{MA}				+	+							
κ_M		-	-	-	-	-	-	-	-	-	-	-
$\theta_{\kappa M}$	+	+	+	+	+	+	+	+	+	+	+	+
$\psi_{\kappa M}$	+											
γ_R	+	+						+				
$\theta_{\gamma R}$	+											
η	+	+	+									
γ_A		-	-	-							-	
$\theta_{\gamma A}$					+							
τ_M	-	-	-	-	-	-	-	-	-	-	-	-
β_T		+	+	+	+	+	+	+	+	+	+	+
$\theta_{\beta T}$						-						
ϵ_T	-	-	-	-	-	-	-	-	-	-	-	-
ϵ_T	+	+	+	+	+	+	+	+	+	+	+	+
τ_T	-	-	-	-				-	-	-	-	-
λ_T		-	-	-	-	-	-	-	-	-	-	-
μ_{TN}	+	+	+									
μ_{TA}	+	+	+	+							+	+

Table C.1: Parameter effects on average lesion size over the course of infection. + denotes the parameter has a significant, positive PRCC value for the lesion size in the 10 day period, whilst - denotes the parameter has a significant, negative PRCC value for the lesion size. Where no symbol is present, the PRCC value is non-significant.

	Days after initial infection											
	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	111-120
S_Q					+	+	+	+				
S_G	+/-	-	-									
λ_R			+	+	+	+	+	+	+	+	+	+
λ_I						-	-					
σ_λ			-	-								
ϕ		+	+	+	+	+	+	+	+			
ξ								-	-			
θ_ξ	+	+										
$\theta_{\tilde{\xi}}$		-										
λ_D	-		+	+	+							
$\theta_{\tau BY}$	+		-									
α_D	-	-	-	-					-	-	-	-
β_D										+	+	
$\theta_{\beta D}$					+	+						
μ_{DI}	-	+	+									
μ_{DM}	+	+										
γ_D	+											
$\theta_{\gamma D}$	-											
τ_D	+											
β_{ML}	+	+	+									
β_{MY}						-	-	-			-	
$\theta_{\beta MY}$		-										
$\theta_{\epsilon MB}$	+											
ϵ_{MT}			-	-								
$\theta_{\epsilon MT}$			+	+	+	+	+	+	+	+	+	+
μ_{MR}			+	+	+	+	+	+	+	+	+	
ζ	-											
κ_M					-	-	-	-	-	-		
$\theta_{\kappa M}$	-		+	+	+	+	+	+	+	+	+	+
$\psi_{\kappa M}$					-	-	-	-	-			
$\theta_{\gamma R}$	+	+	+	+	+	+	+	+				
η	+											
γ_A	+		-	-	-	-	-	-	-	-	-	-
$\theta_{\gamma A}$	-											
τ_M	+	-	-	-	-	-	-	-	-	-	-	-
α_T		+										
β_T	-	-										
$\theta_{\beta T}$	-											
ϵ_T			-	-	-	-	-	-	-	-	-	-
$\theta_{\epsilon T}$	+	+	+	+	+	+	+	+	+	+	+	+
τ_T				-	-	-		-	-	-	-	-
σ_T				+	+	+	+	+	+	+	+	+
λ_T			-	-	-	-	-	-	-	-	-	-
μ_{TN}			+	+	+	+	+					
μ_{TA}	-											+

Table C.2: Parameter effects on number of lesions over the course of infection. + denotes the parameter has a significant, positive PRCC value for the number of lesions in the 10 day period, - denotes the parameter has a significant, negative PRCC value for the number of lesions, and +/- denotes both positive and negative PRCC values exist. Where no symbol is present, the PRCC value is non-significant.

	Days after initial infection											
	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101-110	111-120
S_V	-											
S_Q	+		+	+	+	+	+	+	+	+	+	+
S_G				+								
λ_R	+	+	+	+	+	+	+	+	+	+	+	+
λ_D	+				+	+	+	+	+	+	+	
λ_I	+	+			-	-	-					
σ_λ							-	-	-	-	-	
ϕ	-		+	+	+	+	+					
ξ	+	+	+									
θ_ξ			-									-
σ_ξ	-		-									
τ_{BY}		+	+	+					+	+	+	
$\theta_{\tau BY}$		-	-									
α_D	-	-	-						-	-	-	-
β_D	+		+					+	+	+	+	+
$\theta_{\beta D}$	-	-										
μ_{DI}	+	+	+									
μ_{DM}								-	-			
τ_D	+		-	-	-	-				-		
α_{MY}	+	+										
β_{MY}					-	-	-	-	-	-	-	-
$\theta_{\epsilon MB}$		-	-									
$\theta_{\epsilon MT}$	+	+	+	+	+	+	+	+	+	+	+	+
μ_{MR}	+	+	+	+					+	+	+	+
μ_{MA}				+	+	+						
κ_M		-	-	-	-	-	-	-	-	-	-	-
$\theta_{\kappa M}$	+	+	+	+	+	+	+	+	+	+	+	+
$\psi_{\kappa M}$	+					-						
γ_R	+	+										
$\theta_{\gamma R}$	+			+				+				+
η	+	+	+									
γ_A		-	-	-	-	-	-	-	-	-	-	-
τ_M	-	-	-	-	-	-	-	-	-	-	-	-
β_T			+	+	+	+	+	+	+	+	+	+
$\theta_{\beta T}$						-						
ϵ_T	-	-	-	-	-	-	-	-	-	-	-	-
$\theta_{\epsilon T}$	+	+	+	+	+	+	+	+	+	+	+	+
τ_T	-	-	-	-				-	-	-	-	-
λ_T		-	-	-	-	-	-	-	-	-	-	-
μ_{TN}	+	+	+	+	+							
μ_{TA}	+	+	+								+	+

Table C.3: Parameter effects on total number of bacteria over the course of infection. + denotes the parameter has a significant, positive PRCC value for the total bacteria numbers in the 10 day period, whilst - denotes the parameter has a significant, negative PRCC value for the total bacteria numbers. Where no symbol is present, the PRCC value is non-significant.

Bibliography

[1] J. M. Achkar, J. Chan and A. Casadevall. 'B cells and antibodies in the defense against Mycobacterium tuberculosis infection'. In: *Immunological Reviews* 264.1 (2015), pp. 167–181. ISSN: 1600065X. DOI: 10.1111/imr.12276.

[2] S. Ahmad. 'Pathogenesis, immunology, and diagnosis of latent mycobacterium tuberculosis infection'. In: *Clinical and Developmental Immunology* 2011 (2011). ISSN: 17402522. DOI: 10.1155/2011/814943. arXiv: 814943.

[3] Z. Ahmad, L. G. Klinkenberg, M. L. Pinn, M. M. Fraig, C. A. Peloquin, W. R. Bishai, E. L. Nuermberger, J. H. Grosset and P. C. Karakousis. 'Biphasic Kill Curve of Isoniazid Reveals the Presence of Drug-Tolerant, Not Drug-Resistant, Mycobacterium tuberculosis in the Guinea Pig'. In: *The Journal of Infectious Diseases* 200.7 (2009), pp. 1136–1143. ISSN: 0022-1899. DOI: 10.1086/605605.

[4] A. Alcaïs, C. Fieschi, L. Abel and J.-L. Casanova. 'Tuberculosis in children and adults: two distinct genetic diseases.' In: *The Journal of experimental medicine* 202.12 (2005), pp. 1617–21. ISSN: 0022-1007. DOI: 10.1084/jem.20052302.

[5] N. Alipanah, L. Jarlsberg, C. Miller, N. N. Linh, D. Falzon, E. Jaramillo and P. Nahid. *Adherence interventions and outcomes of tuberculosis treatment: A systematic review and meta-analysis of trials and observational studies*. Vol. 15. 7. 2018, pp. 1–44. ISBN: 1111111111. DOI: 10.1371/journal.pmed.1002595.

[6] D. Alvarez, E. H. Vollmann and U. H. von Andrian. 'Mechanisms and Consequences of Dendritic Cell Migration'. In: *Immunity* 29.3 (2008), pp. 325–342. ISSN: 10747613. DOI: 10.1016/j.immuni.2008.08.006.

[7] E. P. Amaral, E. B. Lasunskaia and M. R. D'Império-Lima. 'Innate immunity in tuberculosis: How the sensing of mycobacteria and tissue damage modulates macrophage death'. In: *Microbes and Infection* 18.1 (2016), pp. 11–20. ISSN: 1769714X. DOI: 10.1016/j.micinf.2015.09.005. arXiv: NIHMS150003.

[8] J. B. Amberson. 'The Significance of Latent Forms of Tuberculosis'. In: *New England Journal of Medicine* 219.15 (Oct. 1938), pp. 572–576. ISSN: 0028-4793. DOI: 10.1056/NEJM193810132191504.

[9] U. H. von Andrian and C. R. Mackay. 'T-Cell Function and Migration — Two Sides of the Same Coin'. In: *New England Journal of Medicine* 343.14 (Oct. 2000). Ed. by I. R. Mackay and F. S. Rosen, pp. 1020–1034. ISSN: 0028-4793. DOI: 10.1056/NEJM200010053431407.

[10] R. Antia, J. C. Koella and V. Perrot. 'Models of the within-host dynamics of persistent mycobacterial infections'. In: *Proc Biol Sci* 263 (1996), pp. 257–263. ISSN: 0962-8452. DOI: 10.1098/rspb.1996.0040[doi].

[11] V. Balasubramanian, E. Wiegeshaus, B. Taylor and D. Smith. 'Pathogenesis of tuberculosis: pathway to apical localization'. In: *Tubercle and Lung Disease* 75.3 (June 1994), pp. 168–178. ISSN: 09628479. DOI: 10.1016/0962-8479(94)90002-7.

[12] J. Banchereau and R. M. Steinman. 'Dendritic cells and the control of immunity.' In: *Nature* 392.March (1998), pp. 245–252. ISSN: 0028-0836. DOI: 10.1038/32588.

[13] A. Bandera et al. 'Molecular Epidemiology Study of Exogenous Reinfection in an Area with a Low Incidence of Tuberculosis'. In: *Journal of Clinical Microbiology* 39.6 (June 2001), pp. 2213–2218. ISSN: 0095-1137. DOI: 10.1128/JCM.39.6.2213-2218.2001.

[14] C. E. Barry, H. I. Boshoff, V. Dartois, T. Dick, S. Ehrt, J. Flynn, D. Schnappinger, R. J. Wilkinson and D. Young. 'The spectrum of latent tuberculosis: rethinking the biology and intervention strategies'. In: *Nature Reviews Microbiology* 7.12 (Dec. 2009), pp. 845–855. ISSN: 1740-1526. DOI: 10.1038/nrmicro2236.

[15] R. J. Basaraba, D. D. Dailey, C. T. McFarland, C. A. Shanley, E. E. Smith, D. N. McMurray and I. M. Orme. 'Lymphadenitis as a major element of disease in the guinea pig model of tuberculosis'. In: *Tuberculosis* 86.5 (2006), pp. 386–394. ISSN: 14729792. DOI: 10.1016/j.tube.2005.11.003.

[16] R. J. Basaraba, E. E. Smith, C. A. Shanley and I. M. Orme. 'Pulmonary lymphatics are primary sites of Mycobacterium tuberculosis infection in guinea pigs infected by aerosol'. In: *Infection and Immunity* 74.9 (2006), pp. 5397–5401. ISSN: 00199567. DOI: 10.1128/IAI.00332-06.

[17] A. L. Bauer, I. B. Hogue, S. Marino and D. E. Kirschner. 'The Effects of HIV – 1 Infection on Latent Tuberculosis'. In: *Math Model Nat Phenom* 3.7 (2008), pp. 229–266.

[18] M. R. Becklake and F. Kauffmann. 'Gender differences in airway behaviour over the human life span'. In: *Thorax* 54.12 (1999), pp. 1119–1138. ISSN: 0040-6376. DOI: 10.1136/thx.54.12.1119.

[19] M. A. Behr. 'Tuberculosis due to Multiple Strains'. In: *American Journal of Respiratory and Critical Care Medicine* 169.5 (Mar. 2004), pp. 554–555. ISSN: 1073-449X. DOI: 10.1164/rccm.2401001. URL: http://www.atsjournals.org/doi/abs/10.1164/rccm.2401001.

[20] M. A. Behr and W. R. Waters. 'Is tuberculosis a lymphatic disease with a pulmonary portal?' In: *The Lancet Infectious Diseases* 14.3 (2014), pp. 250–255. ISSN: 14733099. DOI: 10.1016/S1473-3099(13)70253-6.

[21] S. M. Blower, A. R. Mclean, T. C. Porco, P. M. Small, P. C. Hopewell, M. A. Sanchez and A. R. Moss. 'The intrinsic transmission dynamics of tuberculosis epidemics'. In: *Nature Medicine* 1.8 (Aug. 1995), pp. 815–821. ISSN: 1078-8956. DOI: 10.1038/nm0895-815. arXiv: 9809069v1 [arXiv:gr-qc]. URL: http://www.ncbi.nlm.nih.gov/pubmed/15003161%7B%%7D5Cnhttp://cid.oxfordjournals.org/lookup/doi/10.1093/cid/cir991%7B%%7D5Cnhttp://www.scielo.cl/pdf/udecada/v15n26/art06.pdf%7B%%7D5Cnhttp://www.scopus.com/inward/record.url?eid=2-s2.0-84861150233%7B%5C&%7DpartnerID=tZOtx3y1%20http://www.

[22] K. A. Bodnar, N. V. Serbina and J. L. Flynn. 'Fate of Mycobacterium tuberculosis within murine dendritic cells'. In: *Infection and Immunity* 69.2 (2001), pp. 800–809. ISSN: 00199567. DOI: 10.1128/IAI.69.2.800-809.2001.

[23] P. Bousso. 'T-cell activation by dendritic cells in the lymph node: lessons from the movies.' In: *Nature reviews. Immunology* 8.9 (2008), pp. 675–84. ISSN: 1474-1733. DOI: 10.1038/nri2379.

[24] J. Bowness and A. Taylor. *Anatomy for the FRCA*. Cambridge University Press, July 2019, pp. 2–2. ISBN: 9781108687805. DOI: 10.1017/9781108687805.

[25] R. Bowness. 'Systems Medicine and Infection'. In: *Systems Medicine*. Ed. by U. Schmitz and O. Wolkenhauer. Vol. 1386. Methods in Molecular Biology. New York, NY: Springer New York, 2015, pp. 107–118. ISBN: 978-1-4939-3282-5. DOI: 10.1007/978-1-4939-3283-2.

[26] R. Bowness, M. A. Chaplain, G. G. Powathil and S. H. Gillespie. 'Modelling the effects of bacterial cell state and spatial location on tuberculosis treatment: Insights from a hybrid multiscale cellular automaton model'. In: *Journal of Theoretical Biology* 446 (2018), pp. 87–100. ISSN: 10958541. DOI: 10.1016/j.jtbi.2018.03.006.

[27] P. J. Brennan. 'The Envelope of Mycobacteria'. In: *Annual Review of Biochemistry* 64.1 (2002), pp. 29–63. ISSN: 00664154. DOI: 10.1146/annurev.biochem.64.1.29.

[28] R. J. Brindle, P. P. Nunn, W. Githui, B. W. Allen, S. Gathua and P. Waiyaki. 'Quantitative Bacillary Response to Treatment in HIV-associated Pulmonary Tuberculosis'. In: *American Review of Respiratory Disease* 147.4 (2013), pp. 958–961. ISSN: 0003-0805. DOI: 10.1164/ajrccm/147.4.958.

[29] C. J. Cambier, S. Falkow and L. Ramakrishnan. 'Host evasion and exploitation schemes of Mycobacterium tuberculosis'. In: *Cell* 159.7 (2014), pp. 1497–1509. ISSN: 10974172. DOI: 10.1016/j.cell.2014.11.024.

[30] G. Canetti. *The tubercle bacillus*. New York, 1955.

[31] G. Canetti. 'Exogenous reinfection and pulmonary tuberculosis a study of the pathology.' In: *Tubercle* 31.10 (Oct. 1950), pp. 224–33. ISSN: 0041-3879.

[32] Y. Cao, D. T. Gillespie and L. R. Petzold. 'Efficient step size selection for the tau-leaping simulation method'. In: *Journal of Chemical Physics* 124.4 (2006). ISSN: 00219606. DOI: 10.1063/1.2159468.

[33] S. V. Capuano et al. 'Experimental Mycobacterium tuberculosis infection of cynomolgus macaques closely resembles the various manifestations of human M. tuberculosis infection.' In: *Infection and immunity* 71.10 (2003), pp. 5831–44. ISSN: 0019-9567. DOI: 10.1128/IAI.71.10.5831.

[34] P. J. Cardona. 'A dynamic reinfection hypothesis of latent tuberculosis infection'. In: *Infection* 37.2 (2009), pp. 80–86. ISSN: 03008126. DOI: 10.1007/s15010-008-8087-y.

[35] P. J. Cardona. 'A spotlight on liquefaction: Evidence from clinical settings and experimental models in tuberculosis'. In: *Clinical and Developmental Immunology* 2011 (2011). ISSN: 17402522. DOI: 10.1155/2011/868246.

[36] P. J. Cardona, M. Català, M. Arch, L. Arias, S. Alonso, P. Cardona, D. López, C. Vilaplana and C. Prats. 'Can systems immunology lead tuberculosis eradication?' In: *Current Opinion in Systems Biology* 12 (2018), pp. 53–60. ISSN: 24523100. DOI: 10.1016/j.coisb.2018.10.004. URL: https://doi.org/10.1016/j.coisb.2018.10.004.

[37] P.-J. Cardona. 'New Insights on the Nature of Latent Tuberculosis Infection and its Treatment'. In: *Inflammation & Allergy-Drug Targets* 6.1 (2008), pp. 27–39. ISSN: 18715281. DOI: 10.2174/187152807780077282.

[38] P.-J. Cardona. 'The key role of exudative lesions and their encapsulation: lessons learned from the pathology of human pulmonary tuberculosis'. In: *Frontiers in Microbiology* 6.JUN (June 2015), pp. 1–8. ISSN: 1664-302X. DOI: 10.3389/fmicb.2015.00612.

[39] G. Caron-Lormier, R. W. Humphry, D. A. Bohan, C. Hawes and P. Thorbek. 'Asynchronous and synchronous updating in individual-based models'. In: *Ecological Modelling* 212.3-4 (2008), pp. 522–527. ISSN: 03043800. DOI: 10.1016/j.ecolmodel.2007.10.049.

[40] A. Casadevall and L. anne Pirofski. 'A Reappraisal of Humoral Immunity Based on Mechanisms of Antibody-Mediated Protection Against Intracellular Pathogens'. In: *Advances in Immunology* 91.06 (2006), pp. 1–44. ISSN: 00652776. DOI: 10.1016/S0065-2776(06)91001-3.

[41] Centers for Disease Control. 'Core Curriculum on Tuberculosis : What the Clinician Should Know'. In: *Sixth Edition 2013* (2013), pp. 1–320. ISSN: 0019-9567. DOI: 10.1128/IAI.00057-06.

[42] A. a. Chackerian, J. M. Alt, T. V. Perera, C. C. Dascher and S. M. Behar. 'Dissemination of Mycobacterium tuberculosis Is Influenced by Host Factors and Precedes the Initiation of T-Cell Immunity'. In: *Infection and immunity* 70.8 (2002), pp. 4501–4509. ISSN: 00199567. DOI: 10.1128/IAI.70.8.4501âĂŞ4509.2002.

[43] R. E. Chaisson and N. A. Martinson. 'Tuberculosis in Africa — Combating an HIV-Driven Crisis'. In: *New England Journal of Medicine* 358.11 (Mar. 2008), pp. 1089–1092. DOI: 10.1056/NEJMp0800809.

[44] A. S. Chaves, M. F. Rodrigues, A. M. M. Mattos and H. C. Teixeira. 'Challenging Mycobacterium tuberculosis dormancy mechanisms and their immunodiagnostic potential'. In: *Brazilian Journal of Infectious Diseases* 19.6 (2015), pp. 636–642. ISSN: 16784391. DOI: 10.1016/j.bjid.2015.08.004.

[45] J. Cohen, W. R. Douma, N. H. ten Hacken, M. Oudkerk and D. S. Postma. 'Physiology of the small airways: A gender difference?' In: *Respiratory Medicine* 102.9 (2008), pp. 1264–1271. ISSN: 09546111. DOI: 10.1016/j.rmed.2008.04.007.

[46] V. Colizza and A. Vespignani. 'Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations'. In: *Journal of Theoretical Biology* 251.3 (2008), pp. 450–467. ISSN: 00225193. DOI: 10.1016/j.jtbi.2007.11.028. arXiv: 0706.3647.

[47] H. L. Collins and S. H. E. Kaufmann. 'The many faces of host responses to tuberculosis'. In: *Immunology* 103.1 (2001), pp. 1–9. ISSN: 00192805. DOI: 10.1046/j.1365-2567.2001.01236.x.

[48] C. D. Cook, P. J. Helliesen and S. Agathon. 'Relation Between Mechanics of Respiration, Lung Size and Body Size From Birth to Young Adulthood'. In: *Journal of Applied Physiology* 13.3 (Nov. 1958), pp. 349–352. ISSN: 8750-7587. DOI: 10.1152/jappl.1958.13.3.349.

[49] D. Cornforth, D. G. Green and D. Newth. 'Ordered asynchronous processes in multi-agent systems'. In: *Physica D: Nonlinear Phenomena* 204.1-2 (2005), pp. 70–82. ISSN: 01672789. DOI: 10.1016/j.physd.2005.04.005.

[50] O. Cosivi et al. 'Zoonotic tuberculosis due to Mycobacterium bovis in developing countries.' In: *Emerging infectious diseases* 4.1 (1998), pp. 59–70. ISSN: 1080-6040. DOI: 10.3201/eid0401.980108.

[51] J. E. Cotes, D. J. Chinn and M. R. Miller. *Lung Function*. 6th. Vol. 52. Oxford, UK: Blackwell Publishing Ltd., Mar. 2006, pp. 26–35. ISBN: 9781444312829. DOI: 10.1002/9781444312829.

[52] J. Crofton and D. A. Mitchison. 'Streptomycin Resistance in Pulmonary Tuberculosis'. In: *BMJ* Vol 2 (Dec. 1948), pp. 1009–1015. ISSN: 0959-8138. DOI: 10.1136/bmj.2.4588.1009.

[53] T. M. Daniel. 'The history of tuberculosis'. In: *Respiratory Medicine* 100.11 (2006), pp. 1862–1870. ISSN: 09546111. DOI: 10.1016/j.rmed.2006.08.006.

[54] T. M. Daniel. 'The origins and precolonial epidemiology of tuberculosis in the Americas: Can we figure them out?' In: *International Journal of Tuberculosis and Lung Disease* 4.5 (2000), pp. 395–400. ISSN: 10273719.

[55] A. M. Dannenberg and M. Sugimoto. 'Liquefaction of caseous foci in tuberculosis.' In: *The American review of respiratory disease* 113.3 (1976), pp. 257–9. ISSN: 0003-0805.

[56] A. M. Dannenberg. 'Delayed-type hypersensitivity and cell- mediated immunity in the pathogenesis of tuberculosis'. In: *Immunology Today* 12.7 (1991), pp. 228–233.

[57] A. M. Dannenberg. 'Liquefaction and cavity formation in pulmonary TB: A simple method in rabbit skin to test inhibitors'. In: *Tuberculosis* 89.4 (2009), pp. 243–247. ISSN: 14729792. DOI: 10.1016/j.tube.2009.05.006.

[58] V. Dartois. 'The path of anti-tuberculosis drugs: From blood to lesions to mycobacterial cells'. In: *Nature Reviews Microbiology* 12.3 (2014), pp. 159–167. ISSN: 17401526. DOI: 10.1038/nrmicro3200. arXiv: NIHMS150003.

[59] M. Datta, L. E. Via, W. Chen, J. W. Baish, L. Xu, C. E. Barry and R. K. Jain. 'Mathematical Model of Oxygen Transport in Tuberculosis Granulomas'. In: *Annals of Biomedical Engineering* 44.4 (2016), pp. 863–872. ISSN: 15739686. DOI: 10.1007/s10439-015-1415-3.

[60] A. Davoodi and R. B. Boozarjomehry. 'Developmental model of an automatic production of the human bronchial tree based on L-system'. In: *Computer Methods and Programs in Biomedicine* 132 (2016), pp. 1–10. ISSN: 18727565. DOI: 10.1016/j.cmpb.2016.04.021.

[61] K. Dheda, T. Gumbo, N. R. Gandhi, M. Murray, G. Theron, Z. Udwadia, G. B. Migliori and R. Warren. 'Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis'. In: *The Lancet Respiratory Medicine* 2.4 (Apr. 2014), pp. 321–338. ISSN: 22132600. DOI: 10.1016/S2213-2600(14)70031-1. arXiv: 15334406.

[62] M. Divangahi, M. Chen, H. Gan, D. Desjardins, T. T. Hickman, D. M. Lee, S. Fortune, S. M. Behar and H. G. Remold. 'Mycobacterium tuberculosis evades macrophage defenses by inhibiting plasma membrane repair'. In: *Nature Immunology* 10.8 (2009), pp. 899–906. ISSN: 15292908. DOI: 10.1038/ni.1758.

[63] S. Dobson. *epyc 0.99.1*. 2019. URL: https://pypi.org/project/epyc/ (visited on 17/05/2019).

[64] K. E. Dooley, P. P. Phillips, P. Nahid and M. Hoelscher. 'Challenges in the clinical assessment of novel tuberculosis drugs'. In: *Advanced Drug Delivery Reviews* 102 (2016), pp. 116–122. ISSN: 18728294. DOI: 10.1016/j.addr.2016.01.014.

[65] A. Dorhoi and S. H. E. Kaufmann. 'Pathology and immune reactivity: understanding multidimensionality in pulmonary tuberculosis'. In: *Seminars in Immunopathology* 38.2 (2016), pp. 153–166. ISSN: 18632300. DOI: 10.1007/s00281-015-0531-3.

[66] R. L. Drake, A. W. Vogl and A. W. M. Mitchell. *Gray's Anatomy For Students*. 3rd. Philadelphia, Pennsylvania, USA: Elsevier, 2015, pp. 172–173.

[67] P. T. Elkington and J. S. Friedland. 'Permutations of time and place in tuberculosis'. In: *The Lancet Infectious Diseases* 15.11 (Nov. 2015), pp. 1357–1360. ISSN: 14733099. DOI: 10.1016/S1473-3099(15)00135-8.

[68] A. M. Elliott, S. E. Beming, M. D. Iseman and C. A. Peloquin. 'Failure of drug penetration and acquisition of drug resistance in chronic tuberculous empyema'. In: *Tubercle and Lung Disease* 76.5 (1995), pp. 463–467. ISSN: 09628479. DOI: 10.1016/0962-8479(95)90016-0.

[69] J. J. Ellner. 'Tuberculosis'. In: *Goldman's Cecil Medicine* (2012), pp. 1939–1948. DOI: 10.1016/B978-1-4377-1604-7.00332-8.

[70] R. A. English, Y. Lebovitz and R. B. Griffin. 'Challenges in Clinical Research'. In: *Transforming Clinical Research in the United States*. The National Academic Press, 2010. Chap. 3, pp. 19–36.

[71] S. Y. Eum, J. H. Kong, M. S. Hong, Y. J. Lee, J. H. Kim, S. H. Hwang, S. N. Cho, L. E. Via and C. E. Barry. 'Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB'. In: *Chest* 137.1 (2010), pp. 122–128. ISSN: 19313543. DOI: 10.1378/chest.09-0903. URL: http://dx.doi.org/10.1378/chest.09-0903.

[72] M. Fallahi-Sichani, M. El-Kebir, S. Marino, D. E. Kirschner and J. J. Linderman. 'Multiscale Computational Modeling Reveals a Critical Role for TNF- Receptor 1 Dynamics in Tuberculosis Granuloma Formation'. In: *The Journal of Immunology* 186.6 (2011), pp. 3472–3483. ISSN: 0022-1767. DOI: 10.4049/jimmunol.1003299.

[73] M. Fallahi-Sichani, M. A. Schaller, D. E. Kirschner, S. L. Kunkel and J. J. Linderman. 'Identification of key processes that control tumor necrosis factor availability in a tuberculosis granuloma'. In: *PLoS Computational Biology* 6.5 (2010), pp. 1–19. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1000778.

[74] K. P. Fennelly, J. W. Martyny, K. E. Fulton, I. M. Orme, D. M. Cave and L. B. Heifets. 'Cough-generated Aerosols of Mycobacterium Tuberculosis'. In: *American Journal of Respiratory and Critical Care Medicine* 169.5 (2004), pp. 604–609. ISSN: 1073-449X. DOI: 10.1164/rccm.200308-1101OC.

[75] I. E. A. Flesch and S. H. E. Kaufmann. 'Activation of tuberculostatic macrophage functions by gamma interferon, interleukin-4, and tumor necrosis factor'. In: *Infection and Immunity* 58.8 (1990), pp. 2675–2677. ISSN: 00199567.

[76] K. Floyd, P. Glaziou, A. Zumla and M. Raviglione. 'The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era'. In: *The Lancet Respiratory Medicine* 6.4 (2018), pp. 299–314. ISSN: 22132619. DOI: 10.1016/S2213-2600(18)30057-2.

[77] J. L. Flynn, J. Chan and P. L. Lin. 'Macrophages and control of granulomatous inflammation in tuberculosis.' In: *Mucosal immunology* 4.3 (2011), pp. 271–8. ISSN: 1935-3456. DOI: 10.1038/mi.2011.14. arXiv: NIHMS150003. URL: http://www.ncbi.nlm.nih.gov/pubmed/21430653%7B%%7D5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3311958.

[78] J. L. Flynn and J. Chan. 'Immunology of Tuberculosis'. In: *Annual Review of Immunology* 19.1 (Apr. 2001), pp. 93–129. ISSN: 0732-0582. DOI: 10.1146/annurev.immunol.19.1.93.

[79] W. Fox. 'Whither short-course chemotherapy?' In: *British Journal of Diseases of the Chest* 75.4 (Oct. 1981), pp. 331–357. ISSN: 00070971. DOI: 10.1016/0007-0971(81)90022-X.

[80] W. Fox, G. A. Ellard and D. A. Mitchison. 'Studies on the treatment of tuberculosis undertaken by the British Medical Research Council Tuberculosis Units, 1946-1986, with relevant subsequent publications'. In: *International Journal of Tuberculosis and Lung Disease* 3.10 SUPPL. 2 (1999). ISSN: 10273719.

[81] I. Galvin, G. B. Drummond and M. Nirmalan. 'Distribution of blood flow and ventilation in the lung: Gravity is not the only factor'. In: *British Journal of Anaesthesia* 98.4 (2007), pp. 420–428. ISSN: 00070912. DOI: 10.1093/bja/aem036.

[82] S. Ganguli, D. Gammack and D. E. Kirschner. 'A metapopulation model of granuloma formation in the lung during infection with mycobacterium tuberculosis'. In: *Mathematical Biosciences and Engineering* 2.3 (2005), pp. 535–560. ISSN: 1547-1063 (Print) 1547-1063 (Linking). DOI: 040579197[pii].

[83] N. J. Garton et al. 'Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum'. In: *PLoS Medicine* 5.4 (2008), pp. 0634–0645. ISSN: 15491277. DOI: 10.1371/journal.pmed.0050075.

[84] H. Getahun et al. 'Management of latent Mycobacterium tuberculosis infection: WHO guidelines for low tuberculosis burden countries'. In: *European Respiratory Journal* 46.6 (2015), pp. 1563–1576. ISSN: 13993003. DOI: 10.1183/13993003.01245-2015.

[85] R. Ghanem, D. Higdon and H. Owhadi. 'Introduction to Uncertainty Quantification'. In: *Handbook of Uncertainty Quantification*. Ed. by R. Ghanem, D. Higdon and H. Owhadi. Cham: Springer International Publishing, 2017. Chap. 1, pp. 3–6. ISBN: 978-3-319-12384-4. DOI: 10.1007/978-3-319-12385-1_1.

[86] E. Giacomini, E. Iona, L. Ferroni, M. Miettinen, L. Fattorini, G. Orefici, I. Julkunen and E. M. Coccia. 'Infection of human macrophages and dendritic cells with Mycobacterium tuberculosis induces a differential cytokine gene expression that modulates T cell response.' In: *Journal of immunology (Baltimore, Md. : 1950)* 166.12 (2001), pp. 7033–7041. ISSN: 0022-1767. DOI: 10.4049/jimmunol.166.12.7033.

[87] D. T. Gillespie. 'A general method for numerically simulating the stochastic time evolution of coupled chemical reactions'. In: *Journal of Computational Physics* 22.4 (1976), pp. 403–434. ISSN: 00219991. DOI: 10.1016/0021-9991(76)90041-3.

[88] D. T. Gillespie. 'Exact stochastic simulation of coupled chemical reactions'. In: *The Journal of Physical Chemistry* 81.25 (Dec. 1977), pp. 2340–2361. ISSN: 0022-3654. DOI: 10.1021/j100540a008.

[89] S. H. Gillespie. 'Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective'. In: *Antimicrobial agents and chemotherapy* 46.2 (2002), pp. 267–274. ISSN: 0066-4804. DOI: 10.1128/AAC.46.2.267.

[90] S. H. Gillespie, A. M. Crook, T. D. McHugh, C. M. Mendel, S. K. Meredith, S. R. Murray, F. Pappas, P. P. Phillips and A. J. Nunn. 'Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis'. In: *New England Journal of Medicine* 371.17 (2014), pp. 1577–1587. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1407426.

[91] R. W. Glenny, H. T. Robertson, S. Yamashiro and J. B. Bassingthwaighte. 'Applications of fractal analysis to physiology'. In: *Journal of applied physiology (Bethesda, Md. : 1985)* 70.6 (1991), pp. 2351–2367. ISSN: 8750-7587.

[92] J. E. Gomez and J. D. McKinney. 'M. tuberculosis persistence, latency, and drug tolerance'. In: *Tuberculosis* 84.1-2 (2004), pp. 29–44. ISSN: 14729792. DOI: 10.1016/j.tube.2003.08.003.

[93] M. Gonzalez-Juarrero, T. S. Shim, A. Kipnis, A. P. Junqueira-Kipnis and I. M. Orme. 'Dynamics of Macrophage Cell Populations During Murine Pulmonary Tuberculosis'. In: *The Journal of Immunology* 171.6 (Sept. 2003), pp. 3128–3135. ISSN: 0022-1767. DOI: 10.4049/jimmunol.171.6.3128.

[94] R. A. Goodwin and R. M. Des Prez. 'Apical Localization of Pulmonary Tuberculosis, Chronic Pulmonary Histoplasmosis and Progressive Massive Fibrosis of the Lung'. In: *Chest* 83.5 (1983), pp. 801–805.

[95] J. M. Grange, M. Gandy, P. Farmer and A. Zumla. 'Historical declines in tuberculosis: Nature, nurture and the biosocial model'. In: *International Journal of Tuberculosis and Lung Disease* 5.3 (2001), pp. 208–212. ISSN: 10273719.

[96] J. Grosset. 'Mycobacterium tuberculosis in the Extracellular Compartment: an Underestimated Adversary'. In: *Antimicrobial Agents and Chemotherapy* 47.3 (Mar. 2003), pp. 833–836. ISSN: 0066-4804. DOI: 10.1128/AAC.47.3.833–836.2003.

[97] A. F. Gualtieri and J. P. Hecht. 'Dynamics and Control of Infectious Diseases in Stochastic Metapopulation Models'. In: *Journal of Life Sciences* 5 (2011), pp. 503–508.

[98] E. Guirado and L. S. Schlesinger. 'Modeling the Mycobacterium tuberculosis granuloma - the critical battlefield in host immunity and disease.' In: *Frontiers in immunology* 4.April (2013), p. 98. ISSN: 1664-3224. DOI: 10.3389/fimmu.2013.00098.

[99] U. Gupta and V. Katoch. 'Animal models of tuberculosis'. In: *Tuberculosis* 85.5-6 (Sept. 2005), pp. 277–293. ISSN: 14729792. DOI: 10.1016/j.tube.2005.08.008.

[100] M. C. Gutierrez, S. Brisse, R. Brosch, M. Fabre, B. Omaïs, M. Marmiesse, P. Supply and V. Vincent. 'Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis'. In: *PLoS Pathogens* 1.1 (2005), e5. ISSN: 1553-7366. DOI: 10.1371/journal.ppat.0010005.

[101] M. Gyllenberg and I. Hanski. 'Habitat Deterioration, Habitat Destruction, and Metapopulation Persistence in a Heterogenous Landscape'. In: *Theoretical Population Biology* 52.3 (Dec. 1997), pp. 198–215. ISSN: 00405809. DOI: 10.1006/tpbi.1997.1333.

[102] A. Hagberg. *networkx 2.2*. 2018. URL: https://pypi.org/project/networkx/ (visited on 13/12/2018).

[103] S. Hamis, G. Powathil and M. Chaplain. 'Blackboard to Bedside: A Mathematical Modeling Bottom-Up Approach Toward Personalized Cancer Treatments'. In: *JCO clinical cancer informatics* 3 (2019), pp. 1–11. ISSN: 24734276. DOI: 10.1200/CCI.18.00068.

[104] R. J. H. Hammond, V. O. Baron, K. Oravcova, S. Lipworth and S. H. Gillespie. 'Phenotypic resistance in mycobacteria: is it because I am old or fat that I resist you?' In: *Journal of Antimicrobial Chemotherapy* 70.10 (2015), pp. 2823–2827. ISSN: 0305-7453. DOI: 10.1093/jac/dkv178.

[105] I. Hanski. 'Single-species metapopulation dynamics: concepts, models and observations'. In: *Biological Journal of the Linnean Society* 42.1-2 (1991), pp. 17–38. DOI: j.1095-8312.1991.tb00549.x.

[106] K. L. Helke, J. L. Mankowski and Y. C. Manabe. 'Animal models of cavitation in pulmonary tuberculosis'. In: *Tuberculosis* 86.5 (2006), pp. 337–348. ISSN: 14729792. DOI: 10.1016/j.tube.2005.09.001.

[107] J. C. Helton and F. J. Davis. 'Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems'. In: *Reliability Engineering and System Safety* 81.1 (2003), pp. 23–69. ISSN: 09518320. DOI: 10.1016/S0951-8320(03)00058-9.

[108] J. C. Helton, J. D. Johnson, C. J. Sallaberry and C. B. Storlie. 'Survey of sampling-based methods for uncertainty and sensitivity analysis'. In: *Reliability Engineering and System Safety* 91.10-11 (2006), pp. 1175–1209. ISSN: 09518320. DOI: 10.1016/j.ress.2005.11.017.

[109] J. Herman and W. Usher. *SALib: An open-source Python library for Sensitivity Analysis*. Tech. rep. 9. Jan. 2017, p. 97. DOI: 10.21105/joss.00097.

[110] M. Herrera, P. Bosch, M. Nájera and X. Aguilera. 'Modeling the spread of tuberculosis in semiclosed communities'. In: *Computational and Mathematical Methods in Medicine* 2013 (2013). ISSN: 1748670X. DOI: 10.1155/2013/648291.

[111] R. I. Hickson, G. N. Mercer and K. M. Lokuge. 'A metapopulation model of tuberculosis transmission with a case study from high to low burden areas'. In: *PLoS ONE* 7.4 (2012). ISSN: 19326203. DOI: 10.1371/journal.pone.0034411.

[112] K. Horsfield. 'The structure of the tracheobronchial tree'. In: *Scientific Foundation of Respiratory Medicine*. Ed. by J. G. Scadding, G. Cumming and W. M. Thurlbeck. London: Heinemann, 1981. Chap. 6, pp. 54–70.

[113] K. Horsfield and G. Cumming. 'Morphology of the bronchial tree in man'. In: *Journal of Applied Physiology* 24.3 (1968), pp. 373–383.

[114] K. Horsfield, W. I. Gordon, W. Kemp and S. Phillips. 'Growth of the bronchial tree in man.' In: *Thorax* 42.5 (1987), pp. 383–8. ISSN: 0040-6376. DOI: 10.1136/thx.42.5.383.

[115] R. M. G. J. Houben and P. J. Dodd. 'The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling'. In: *PLOS Medicine* 13.10 (2016), e1002152. DOI: 10.1371/journal.pmed.1002152.

[116] V. N. Houk, D. C. Kent, K. Sorensen and J. H. Baker. 'The Eradication of Tuberculosis Infection by Isoniazid Chemoprophylaxis'. In: *Archives of Environmental Health* 16.1 (1968), pp. 46–50. ISSN: 00039896. DOI: 10.1080/00039896.1968.10665013.

[117] Y. Hu, J. A. Mangan, J. Dhillon, K. M. Sole, D. A. Mitchison, P. D. Butcher and A. R. M. Coates. 'Detection of mRNA Transcripts and Active Transcription in Persistent Mycobacterium tuberculosis Induced by Exposure to Rifampin or Pyrazinamide'. In: *Journal of Bacteriology* 182.22 (Nov. 2000), pp. 6358–6365. ISSN: 0021-9193. DOI: 10.1128/JB.182.22.6358-6365.2000.

[118] P. Hudelson. 'Gender differentials in tuberculosis: The role of socio-economic and cultural factors'. In: *Tubercle and Lung Disease* 77.5 (1996), pp. 391–400. ISSN: 09628479. DOI: 10.1016/S0962-8479(96)90110-0.

[119] I. R. Humphreys, G. R. Stewart, D. J. Turner, J. Patel, D. Karamanou, R. J. Snelgrove and D. B. Young. 'A role for dendritic cells in the dissemination of mycobacterial infection'. In: *Microbes and Infection* 8.5 (2006), pp. 1339–1346. ISSN: 12864579. DOI: 10.1016/j.micinf.2005.12.023.

[120] R. L. Hunter. 'On the pathogenesis of post primary tuberculosis: The role of bronchial obstruction in the pathogenesis of cavities'. In: *Tuberculosis* 91.SUPPL. 1 (2011), S6–S10. ISSN: 14729792. DOI: 10.1016/j.tube.2011.10.003.

[121] R. L. Hunter. 'Pathology of post primary tuberculosis of the lung: An illustrated critical review'. In: *Tuberculosis* 91.6 (2011), pp. 497–509. ISSN: 14729792. DOI: 10.1016/j.tube.2011.03.007. arXiv: NIHMS150003.

[122] R. L. Hunter. 'Tuberculosis as a three-act play: A new paradigm for the pathogenesis of pulmonary tuberculosis'. In: *Tuberculosis* 97 (2016), pp. 8–17. ISSN: 1873281X. DOI: 10.1016/j.tube.2015.11.010.

[123] R. L. Hunter, J. K. Actor, S. A. Hwang, V. Karev and C. Jagannath. 'Pathogenesis of post primary tuberculosis: Immunity and hypersensitivity in the development of cavities'. In: *Annals of Clinical and Laboratory Science* 44.4 (2014), pp. 365–387. ISSN: 15508080.

[124] R. L. Hunter, C. Jagannath and J. K. Actor. 'Pathology of postprimary tuberculosis in humans and mice: Contradiction of long-held beliefs'. In: *Tuberculosis* 87.4 (2007), pp. 267–278. ISSN: 14729792. DOI: 10.1016/j.tube.2006.11.003.

[125] R. L. Hunter, M. R. Olsen, C. Jagannath and J. K. Actor. 'Multiple roles of cord factor in the pathogenesis of primary, secondary, and cavitary tuberculosis, including a revised description of the pathology of secondary disease.' In: *Annals of clinical and laboratory science* 36.4 (2006), pp. 371–86. ISSN: 0091-7370.

[126] M. Jassal and W. R. Bishai. 'Extensively drug-resistant tuberculosis'. In: *The Lancet Infectious Diseases* 9.1 (2009), pp. 19–30. ISSN: 14733099. DOI: 10.1016/S1473-3099(08)70260-3.

[127] A. Jindani et al. 'High-Dose Rifapentine with Moxifloxacin for Pulmonary Tuberculosis'. In: *New England Journal of Medicine* 371.17 (Oct. 2014), pp. 1599–1608. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1314210.

[128] P. C. Karakousis. 'Latent Tuberculosis Infection: Myths, Models, and Molecular Mechanisms.' In: *Microbiology and molecular biology reviews : MMBR* 78.3 (2014), pp. 343–371. ISSN: 1092-2172. DOI: 10.1128/MMBR.00010-14.

[129] J. Karumbi and P. Garner. 'Directly observed therapy for treating tuberculosis'. In: *Cochrane Database of Systematic Reviews* 4 (May 2015), pp. 1–56. ISSN: 14651858. DOI: 10.1002/14651858.CD003343.pub4.

[130] A. C. Kirby, M. C. Coles and P. M. Kaye. 'Alveolar Macrophages Transport Pathogens to Lung Draining Lymph Nodes'. In: *The Journal of Immunology* 183.3 (Aug. 2009), pp. 1983–1989. ISSN: 0022-1767. DOI: 10.4049/jimmunol.0901089.

[131] D. Kirschner. 'Dynamics of co-infection with M. Tuberculosis and HIV-1.' In: *Theoretical population biology* 55.1 (1999), pp. 94–109. ISSN: 0040-5809. DOI: 10.1006/tpbi.1998.1382.

[132] D. Kirschner, E. Pienaar, S. Marino and J. J. Linderman. 'A review of computational and mathematical modeling contributions to our understanding of Mycobacterium tuberculosis within-host infection and treatment'. In: *Current Opinion in Systems Biology* 3 (2017), pp. 170–185. ISSN: 24523100. DOI: 10.1016/j.coisb.2017.05.014.

[133] J. M. Ko, H. J. Park and C. H. Kim. 'Clinicoradiologic evidence of pulmonary lymphatic spread in adult patients with tuberculosis'. In: *American Journal of Roentgenology* 204.1 (2015), pp. 38–43. ISSN: 15463141. DOI: 10.2214/AJR-14.12908.

[134] G. H. Kramer, K. Capello, B. Bearrs, A. Lauzon and L. Normandeau. 'Linear dimensions and volumes of human lungs obtained from CT images'. In: *Health Physics* 102.4 (2012), pp. 378–383. ISSN: 00179078. DOI: 10.1097/HP.0b013e31823a13f1.

[135] M. F. Krummel, F. Bartumeus and A. Gérard. 'T cell migration, search strategies and mechanisms'. In: *Nature Reviews Immunology* 16.3 (Mar. 2016), pp. 193–201. ISSN: 1474-1733. DOI: 10.1038/nri.2015.16.

[136] J. Lee, M. Hartman and H. Kornfeld. 'Macrophage apoptosis in tuberculosis'. In: *Yonsei Medical Journal* 50.1 (2009), pp. 1–11. ISSN: 05135796. DOI: 10.3349/ymj.2009.50.1.1.

[137] A. Lenaerts, C. E. Barry and V. Dartois. 'Heterogeneity in tuberculosis pathology, microenvironments and therapeutic responses'. In: *Immunological Reviews* 264.1 (2015), pp. 288–307. ISSN: 1600065X. DOI: 10.1111/imr.12252.

[138] K. O. Leslie and M. R. Wick. *Lung Anatomy*. Third Edit. Elsevier Inc., 2017, pp. 1–13. ISBN: 9781455711796. DOI: 10.1016/B978-0-323-44284-8.00001-6.

[139] R. Levins. 'Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control'. In: *Bulletin of the Entomological Society of America* 15.3 (Sept. 1969), pp. 237–240. ISSN: 0013-8754. DOI: 10.1093/besa/15.3.237.

[140] T. Lietman and S. M. Blower. 'Potential Impact of Tuberculosis Vaccines as Epidemic Control Agents'. In: *Clinical Infectious Diseases* 30.Supplement 3 (2000), S316–S322. ISSN: 1058-4838. DOI: 10.1086/313881.

[141] P. L. Lin and J. L. Flynn. 'CD8 T cells and Mycobacterium tuberculosis infection'. In: *Seminars in Immunopathology* 37.3 (May 2015), pp. 239–249. ISSN: 1863-2297. DOI: 10.1007/s00281-015-0490-8.

[142] S. Lipworth, R. Hammond, V. Baron, Y. Hu, A. Coates and S. Gillespie. 'Defining dormancy in mycobacterial disease'. In: *Tuberculosis* 99 (July 2016), pp. 131–142. ISSN: 14729792. DOI: 10.1016/j.tube.2016.05.006.

[143] A. L. Lloyd and R. M. May. 'Spatial heterogeneity in epidemic models.' In: *Journal of theoretical biology* 179.1 (1996), pp. 1–11. ISSN: 0022-5193. DOI: 10.1006/jtbi.1996.0042.

[144] R. O. Loebel, E. Shorr and H. B. Richardson. 'The Influence of Foodstuffs upon the Respiratory Metabolism and Growth of Human Tubercle Bacilli.' In: *Journal of bacteriology* 26.2 (Aug. 1933), pp. 139–66. ISSN: 0021-9193.

[145] D. M. Lowe, P. S. Redford, R. J. Wilkinson, A. O'Garra and A. R. Martineau. 'Neutrophils in tuberculosis: friend or foe?' In: *Trends in Immunology* 33.1 (Jan. 2012), pp. 14–25. ISSN: 14714906. DOI: 10.1016/j.it.2011.10.003. URL: http://dx.doi.org/10.1016/j.it.2011.10.003%20https://linkinghub.elsevier.com/retrieve/pii/S1471490611001827.

[146] P. J. Maglione and J. Chan. 'How B cells shape the immune response against Mycobacterium tuberculosis'. In: *European Journal of Immunology* 39.3 (2009), pp. 676–686. ISSN: 00142980. DOI: 10.1002/eji.200839148.

[147] G. Magombedze, W. Garira and E. Mwenje. 'Modelling the human immune response mechanisms to mycobacterium tuberculosis infection in the lungs'. In: *Math Biosci Eng* 3.3 (2006), pp. 661–682. ISSN: 1547-1063 (Print) 1547-1063 (Linking). DOI: 10.3934/mbe.2006.3.661.

[148] G. Magombedze and N. Mulder. 'A mathematical representation of the development of Mycobacterium tuberculosis active, latent and dormant stages'. In: *Journal of Theoretical Biology* 292 (2012), pp. 44–59. ISSN: 00225193. DOI: 10.1016/j.jtbi.2011.09.025.

[149] B. B. Mandelbrot. 'Trees and the Diameter Exponent'. In: *The Fractal Geometry of Nature*. New York, NY: W H Freeman and Company, 1977. Chap. 17, pp. 157–165.

[150] S. Marino, I. B. Hogue, C. J. Ray and D. E. Kirschner. 'A methodology for performing global uncertainty and sensitivity analysis in systems biology'. In: *J Theor Biol* 254.1 (2008), pp. 178–196. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2008.04.011.

[151] S. Marino, S. Pawar, C. L. Fuller, T. A. Reinhart, J. L. Flynn and D. E. Kirschner. 'Dendritic Cell Trafficking and Antigen Presentation in the Human Immune Response to Mycobacterium tuberculosis'. In: *The Journal of Immunology* 173.1 (2004), pp. 494–506. ISSN: 0022-1767. DOI: 10.4049/jimmunol.173.1.494.

[152] S. Marino, M. El-Kebir and D. Kirschner. 'A hybrid multi-compartment model of granuloma formation and T cell priming in Tuberculosis'. In: *Journal of Theoretical Biology* 280.1 (2011), pp. 50–62. ISSN: 00225193. DOI: 10.1016/j.jtbi.2011.03.022.

[153] S. Marino and D. E. Kirschner. 'The human immune response to Mycobacterium tuberculosis in lung and lymph node.' In: *Journal*

of Theoretical Biology 227.4 (2004), pp. 463–486. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2003.11.023.

[154] S. Marino, A. Myers, J. L. Flynn and D. E. Kirschner. 'TNF and IL-10 are major factors in modulation of the phagocytic cell environment in lung and lymph node in tuberculosis: A next-generation two-compartmental model'. In: *Journal of Theoretical Biology* 265.4 (2010), pp. 586–598. ISSN: 00225193. DOI: 10.1016/j.jtbi.2010.05.012.

[155] D. Masopust and J. M. Schenkel. 'The integration of T cell migration, differentiation and function'. In: *Nat Rev Immunol* 13.5 (2013), pp. 309–320. ISSN: 1474-1741 (Electronic) 1474-1733 (Linking). DOI: 10.1038/nri3442.

[156] L. Mauri and R. B. D'Agostino. 'Challenges in the design and interpretation of noninferiority trials'. In: *New England Journal of Medicine* 377.14 (2017), pp. 1357–1367. ISSN: 15334406. DOI: 10.1056/NEJMra1510063.

[157] R. K. McCormack and L. J. S. Allen. 'Multi-patch deterministic and stochastic models for wildlife diseases'. In: *Journal of Biological Dynamics* 1.1 (Jan. 2007), pp. 63–85. ISSN: 1751-3758. DOI: 10.1080/17513750601032711.

[158] W. McDermott. 'Antimicrobial therapy of pulmonary tuberculosis.' In: *Bulletin of the World Health Organization* 23.4-5 (1960), pp. 427–61. ISSN: 0042-9686.

[159] E. M. Medlar. 'Necropsy studies of human pulmonary tuberculosis'. In: *Am Rev Tuberc* 71.2 (1955), pp. 29–55.

[160] E. Medlar. 'Pathogenetic concepts of tuberculosis'. In: *The American Journal of Medicine* 9.5 (Nov. 1950), pp. 611–622. ISSN: 00029343. DOI: 10.1016/0002-9343(50)90211-7. URL: https://linkinghub.elsevier.com/retrieve/pii/0002934350902117.

[161] L. S. Meena and T. Rajni. 'Survival mechanisms of pathogenic Mycobacterium tuberculosis H 37Rv'. In: *FEBS Journal* 277.11 (2010), pp. 2416–2427. ISSN: 1742464X. DOI: 10.1111/j.1742-4658.2010.07666.x.

[162] C. S. Merle et al. 'A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis'. In: *New England Journal of Medicine* 371.17 (Oct. 2014), pp. 1588–1598. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1315817.

[163] A. Mert et al. 'Miliary tuberculosis: Clinical manifestations, diagnosis and outcome in 38 adults'. In: *Respirology* 6.3 (2001), pp. 217–224. ISSN: 13237799. DOI: 10.1046/j.1440-1843.2001.00328.x.

[164] A. Mihret. 'The role of dendritic cells in Mycobacterium tuberculosis infection.' In: *Virulence* 3.7 (2012), pp. 654–9. ISSN: 2150-5608. DOI: 10.4161/viru.22586.

[165] H. J. Milburn. 'Primary tuberculosis'. In: *Current Opinion in Pulmonary Medicine* 7.3 (2001), pp. 133–141. ISSN: 10705287. DOI: 10.1097/00063198-200105000-00004.

[166] A. H. Mines. *Respiratory Physiology*. Second. New York, NY: Raven Press, 1986. ISBN: 0-88167-187-8.

[167] R. N. Miranda, J. D. Khoury and L. J. Medeiros. 'Mycobacterium Tuberculosis Lymphadenitis'. In: *Atlas of Lymph Node Pathology*. Vol. 53. New York, NY: Springer New York, 2013, pp. 25–28. ISBN: 9788578110796. DOI: 10.1007/978-1-4614-7959-8_6. arXiv: arXiv:1011.1669v3.

[168] B. B. Mishra et al. 'Nitric oxide prevents a pathogen-permissive granulocytic inflammation during tuberculosis'. In: *Nature Microbiology* 2.May (2017), pp. 1–11. ISSN: 20585276. DOI: 10.1038/nmicrobiol.2017.72.

[169] D. M. Mosser and J. P. Edwards. 'Exploring the full spectrum of macrophage activation'. In: *Nature Reviews Immunology* 8.12 (Dec. 2008), pp. 958–969. ISSN: 1474-1733. DOI: 10.1038/nri2448. arXiv: NIHMS150003.

[170] D. P. Moualeu, M. Weiser, R. Ehrig and P. Deuflhard. 'Optimal control for a tuberculosis model with undetected cases in Cameroon'. In: *Communications in Nonlinear Science and Numerical Simulation*

20.3 (2015), pp. 986–1003. ISSN: 10075704. DOI: 10.1016/j.cnsns.2014.06.037.

[171] G. V. Mukamolova, O. Turapov, J. Malkin, G. Woltmann and M. R. Barer. 'Resuscitation-promoting factors reveal an occult population of tubercle bacilli in sputum'. In: *American Journal of Respiratory and Critical Care Medicine* 181.2 (2010), pp. 174–180. ISSN: 1073449X. DOI: 10.1164/rccm.200905-0661OC.

[172] M. E. Murphy et al. 'Gender differences in tuberculosis treatment outcomes: a post hoc analysis of the REMoxTB study'. In: *BMC Medicine* 16.1 (Dec. 2018), p. 189. ISSN: 1741-7015. DOI: 10.1186/s12916-018-1169-5.

[173] J. F. Murray. 'Bill Dock and the Location of Pulmonary Tuberculosis: How Bed Rest Might Have Helped Consumption'. In: *American Journal of Respiratory and Critical Care Medicine* 168.9 (2003), pp. 1029–1033. ISSN: 1073449X. DOI: 10.1164/rccm.200307-1016OE.

[174] A. J. Myers, S. Marino, D. E. Kirschner and J. L. Flynn. 'Inoculation dose of Mycobacterium tuberculosis does not influence priming of T cell responses in lymph nodes.' In: *Journal of immunology (Baltimore, Md. : 1950)* 190.9 (2013), pp. 4707–16. ISSN: 1550-6606. DOI: 10.4049/jimmunol.1203465.

[175] J. B. Nachega and R. E. Chaisson. 'Tuberculosis Drug Resistance: A Global Threat'. In: *Clinical Infectious Diseases* 36.Supplement_1 (2003), S24–S30. ISSN: 1537-6591. DOI: 10.1086/344657.

[176] H. Ndlovu and M. J. Marakalala. 'Granulomas and inflammation: Host-directed therapies for tuberculosis'. In: *Frontiers in Immunology* 7.OCT (2016). ISSN: 16643224. DOI: 10.3389/fimmu.2016.00434.

[177] T. R. Nelson and D. K. Manchester. 'Modeling of lung morphogenesis using fractal geometries'. In: *IEEE Transactions on Medical Imaging* 7.4 (1988), pp. 321–327. ISSN: 02780062. DOI: 10.1109/42.14515.

[178] A. G. Nerlich, C. J. Haas, A. Zink, U. Szeimies and H. G. Hagedorn. 'Molecular evidence for tuberculosis in an ancient Egyptian mummy'. In: *The Lancet* 350.9088 (Nov. 1997), p. 1404. ISSN: 01406736. DOI: 10.1016/S0140-6736(05)65185-9.

[179] O. Neyrolles and L. Quintana-Murci. 'Sexual inequality in tuberculosis'. In: *PLoS Medicine* 6.12 (2009). ISSN: 15491277. DOI: 10.1371/journal.pmed.1000199.

[180] C. Nimmo, M. Lipman, P. P. Phillips, T. McHugh, A. Nunn and I. Abubakar. 'Shortening treatment of tuberculosis: Lessons from fluoroquinolone trials'. In: *The Lancet Infectious Diseases* 15.2 (2015), pp. 141–143. ISSN: 14744457. DOI: 10.1016/S1473-3099(14)70885-0.

[181] E. Nuermberger, W. Bishai and J. Grosset. 'Latent Tuberculosis Infection.' In: *Semin. Respir. Crit. Care Med.* 25.3 (2004), pp. 317–336.

[182] M. Ochs, J. R. Nyengaard, A. Jung, L. Knudsen, M. Voigt, T. Wahlers, J. Richter and H. J. G. Gundersen. 'The Number of Alveoli in the Human Lung'. In: *American Journal of Respiratory and Critical Care Medicine* 169.1 (2003), pp. 120–124. ISSN: 1073-449X. DOI: 10.1164/rccm.200308-1107oc.

[183] M. Oddo, T. Renno, A. Attinger, T. Bakker, H. R. MacDonald and P. R. Meylan. 'Fas ligand-induced apoptosis of infected human macrophages reduces the viability of intracellular Mycobacterium tuberculosis.' In: *Journal of immunology (Baltimore, Md. : 1950)* 160.11 (1998), pp. 5448–54. ISSN: 0022-1767.

[184] P. L. Olliaro and M. Vaillant. 'Designing noninferiority tuberculosis treatment trials: Identifying practical advantages for drug regimens with acceptable effectiveness'. In: *PLOS Medicine* 16.7 (July 2019). ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1002850. URL: http://dx.plos.org/10.1371/journal.pmed.1002850.

[185] I. M. Orme and R. J. Basaraba. 'The formation of the granuloma in tuberculosis infection'. In: *Seminars in Immunology* 26.6 (2014),

pp. 601–609. ISSN: 10963618. DOI: 10.1016/j.smim.2014.09.009.

[186] L. P. Ormerod. 'Multidrug-resistant tuberculosis (MDR-TB): Epidemiology, prevention and treatment'. In: *British Medical Bulletin* 73-74 (2005), pp. 17–24. ISSN: 00071420. DOI: 10.1093/bmb/ldh047.

[187] F. M. R. Perrin, N. Woodward, P. P. J. Phillips, T. D. McHugh, A. J. Nunn, M. C. I. Lipman and S. H. Gillespie. 'Radiological cavitation, sputum mycobacterial load and treatment response in pulmonary tuberculosis'. In: *International Journal of Tuberculosis and Lung Disease* 14.12 (2010), pp. 1596–1602. ISSN: 10273719. DOI: 144/9/650[pii].

[188] F. M. R. Perrin, M. C. I. Lipman, T. D. McHugh and S. H. Gillespie. 'Biomarkers of treatment response in clinical trials of novel anti-tuberculosis agents'. In: *Lancet Infectious Diseases* 7.7 (2007), pp. 481–490. ISSN: 14733099. DOI: 10.1016/S1473-3099(07)70112-3.

[189] P. P. J. Phillips, C. M. Mendel, D. A. Burger, A. Crook, A. J. Nunn, R. Dawson, A. H. Diacon and S. H. Gillespie. 'Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials'. In: *BMC Medicine* 14.1 (2016), p. 19. ISSN: 1741-7015. DOI: 10.1186/s12916-016-0565-y.

[190] E. Pienaar, N. A. Cilfone, P. L. Lin, V. Dartois, J. T. Mattila, J. R. Butler, J. L. Flynn, D. E. Kirschner and J. J. Linderman. 'A computational tool integrating host immunity with antibiotic dynamics to study tuberculosis treatment'. In: *Journal of Theoretical Biology* 367 (2015), pp. 166–179. ISSN: 00225193. DOI: 10.1016/j.jtbi.2014.11.021.

[191] K. E. Pinkerton, L. S. Van Winkle, C. G. Plopper, S. Smiley-Jewell, E. C. Covarrubias and J. T. McBride. 'Architecture of the Tracheobronchial Tree'. In: *Comparative Biology of the Normal Lung*. Elsevier, 2015. Chap. 4, pp. 33–51. ISBN: 9780124045774. DOI: 10.1016/B978-0-12-404577-4.00004-7.

[192] M. J. Pitcher, R. Bowness, S. Dobson, R. Eftimie and S. H. Gillespie. 'Modelling the effects of environmental heterogeneity within the lung on the tuberculosis life-cycle'. In: *bioRxiv* (2019). DOI: 10.1101/2019.12.12.871269.

[193] M. J. Pitcher, R. Bowness, S. Dobson and S. H. Gillespie. 'A spatially heterogeneous network-based metapopulation software model applied to the simulation of a pulmonary tuberculosis infection'. In: *Applied Network Science* 3.1 (Dec. 2018), p. 33. ISSN: 2364-8228. DOI: 10.1007/s41109-018-0091-2.

[194] M. Pitcher, R. Bowness, S. Dobson and S. Gillespie. 'A Network-Based Metapopulation Model to Simulate a Pulmonary Tuberculosis Infection'. In: *The 6th International Conference on Complex Networks and Their Applications BOOK OF ABSTRACTS*. Lyon, 2017, pp. 61–63.

[195] P. C. Pratt. 'Pathology of tuberculosis'. In: *Seminars in Roentgenology* 14.3 (July 1979), pp. 196–203. ISSN: 0037198X. DOI: 10.1016/0037-198X(79)90006-3.

[196] B. Prideaux et al. 'The association between sterilizing activity and drug distribution into tuberculosis lesions'. In: *Nature Medicine* 21.10 (2015). ISSN: 1078-8956. DOI: 10.1038/nm.3937.

[197] J. Raffetseder, E. Pienaar, R. Blomgran, D. Eklund, V. P. Brodin, H. Andersson, A. Welin and M. Lerm. 'Replication rates of mycobacterium tuberculosis in human macrophages do not correlate with mycobacterial antibiotic susceptibility'. In: *PLoS ONE* 9.11 (2014), pp. 1–10. ISSN: 19326203. DOI: 10.1371/journal.pone.0112426.

[198] S. Raghunandanan, L. Jose and R. A. Kumar. 'Dormant Mycobacterium tuberculosis converts isoniazid to the active drug in a Wayne's model of dormancy'. In: *Journal of Antibiotics* 71.11 (2018), pp. 939–949. ISSN: 18811469. DOI: 10.1038/s41429-018-0098-z.

[199] L. Ramakrishnan. 'Revisiting the role of the granuloma in tuberculosis.' In: *Nature reviews. Immunology* 12.5 (2012), pp. 352–66. ISSN: 1474-1741. DOI: 10.1038/nri3211.

[200] J. C. J. Ray, J. L. Flynn and D. E. Kirschner. 'Synergy between individual TNF-dependent functions determines granuloma performance for controlling Mycobacterium tuberculosis infection.' In: *Journal of immunology (Baltimore, Md. : 1950)* 182.14 (2009), pp. 3706–3717. ISSN: 1550-6606. DOI: 10.4049/jimmunol.0802297.

[201] A. van Rie, R. Warren, M. Richardson, T. C. Victor, R. P. Gie, D. A. Enarson, N. Beyers and P. D. van Helden. 'Exogenous Reinfection as a Cause of Recurrent Tuberculosis after Curative Treatment'. In: *New England Journal of Medicine* 341.16 (Oct. 1999), pp. 1174–1179. ISSN: 0028-4793. DOI: 10.1056/NEJM199910143411602.

[202] L. L. Roberts and C. M. Robinson. 'Mycobacterium tuberculosis infection of human dendritic cells decreases integrin expression, adhesion and migration to chemokines'. In: *Immunology* 141.1 (2014), pp. 39–51. ISSN: 00192805. DOI: 10.1111/imm.12164.

[203] J. a. Rosati Rowe, R. Burton, G. McGregor, R. McCauley, W. Tang and R. Spencer. 'Development of a three-dimensional model of the human respiratory system for dosimetric use.' In: *Theoretical biology & medical modelling* 10.1 (2013), p. 28. ISSN: 1742-4682. DOI: 10.1186/1742-4682-10-28.

[204] E. J. Rubin. 'The Granuloma in Tuberculosis — Friend or Foe?' In: *New England Journal of Medicine* 360.23 (2009), pp. 2471–2473. ISSN: 0028-4793. DOI: 10.1056/nejmcibr0902539.

[205] D. G. Russell, P.-J. Cardona, M.-J. Kim, S. Allain and F. Altare. 'Foamy macrophages and the progression of the human tuberculosis granuloma'. In: *Nature Immunology* 10.9 (Sept. 2009), pp. 943–948. ISSN: 1529-2908. DOI: 10.1038/ni.1781.

[206] D. G. Russell, H. C. Mwandumba and E. E. Rhoades. 'Mycobacterium and the coat of many lipids'. In: *Journal of Cell Biology* 158.3 (2002), pp. 421–426. ISSN: 00219525. DOI: 10.1083/jcb.200205034.

[207] N. Sahay, M. Ierapetritou and J. Wassick. 'Synchronous and asynchronous decision making strategies in supply chains'. In: *Computers and Chemical Engineering* 71 (2014), pp. 116–129. ISSN: 00981354. DOI: 10.1016/j.compchemeng.2014.07.005.

[208] A. Saltelli, S. Tarantola and K. P.-S. Chan. 'A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output'. In: *Technometrics* 41.1 (Feb. 1999), p. 39. ISSN: 00401706. DOI: 10.2307/1270993.

[209] A. Saltelli. 'Sensitivity analysis for importance assessment'. In: *Risk Analysis* 22.3 (2002), pp. 579–590. ISSN: 02724332. DOI: 10.1111/0272-4332.00040.

[210] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li and Q. Wu. 'Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices'. In: *Environmental Modelling and Software* 114.March 2018 (2019), pp. 29–39. ISSN: 13648152. DOI: 10.1016/j.envsoft.2019.01.012.

[211] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola. 'Introduction to Sensitivity Analysis'. In: *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley & Sons, Ltd, Dec. 2008. Chap. 1, pp. 1–51. ISBN: 9780470725184. DOI: 10.1002/9780470725184.ch1.

[212] B. Schönfisch and A. De Roos. 'Synchronous and asynchronous updating in cellular automata'. In: *BioSystems* 51.3 (1999), pp. 123–143. ISSN: 03032647. DOI: 10.1016/S0303-2647(99)00025-8.

[213] J. L. Segovia-Juarez, S. Ganguli and D. Kirschner. 'Identifying control mechanisms of granuloma formation during M. tuberculosis infection using an agent-based model'. In: *Journal of Theoretical Biology* 231.3 (2004), pp. 357–376. ISSN: 00225193. DOI: 10.1016/j.jtbi.2004.06.031.

[214] C. L. Sershen, S. J. Plimpton and E. E. May. 'Oxygen Modulates the Effectiveness of Granuloma Mediated Host Response to Mycobacterium tuberculosis: A Multiscale Computational Biology

Approach'. In: *Frontiers in Cellular and Infection Microbiology* 6.February (2016), pp. 1–25. ISSN: 2235-2988. DOI: 10.3389/fcimb.2016.00006.

[215] T. Shahzad and M. Irfan. 'Endobronchial tuberculosis—a review'. In: *Journal of Thoracic Disease* 8.12 (Dec. 2016), pp. 3797–3802. ISSN: 20721439. DOI: 10.21037/jtd.2016.12.73. URL: http://jtd.amegroups.com/article/view/11238/9716.

[216] M. O. Shleeva, Y. K. Kudykina, G. N. Vostroknutova, N. E. Suzina, A. L. Mulyukin and A. S. Kaprelyants. 'Dormant ovoid cells of Mycobacterium tuberculosis are formed in response to gradual external acidification'. In: *Tuberculosis* 91.2 (2011), pp. 146–154. ISSN: 14729792. DOI: 10.1016/j.tube.2010.12.006.

[217] M. Silva Miranda, A. Breiman, S. Allain, F. Deknuydt and F. Altare. 'The Tuberculous Granuloma: An Unsuccessful Host Defence Mechanism Providing a Safety Shelter for the Bacteria?' In: *Clinical and Developmental Immunology* 2012 (2012), pp. 1–14. ISSN: 1740-2522. DOI: 10.1155/2012/139127.

[218] K. C. Stone, R. R. Mercer, P. Gehr, B. Stockstill and J. D. Crapo. 'Allometric relationships of cell numbers and size in the mammalian lung.' In: *American journal of respiratory cell and molecular biology* 6.2 (1992), pp. 235–243. ISSN: 10441549. DOI: 10.1165/ajrcmb/6.2.235.

[219] D. Sud, C. Bigbee, J. L. Flynn and D. E. Kirschner. 'Contribution of CD8+ T Cells to Control of Mycobacterium tuberculosis Infection'. In: *The Journal of Immunology* 176.7 (2006), pp. 4296–4314. ISSN: 0022-1767. DOI: 10.4049/jimmunol.176.7.4296.

[220] A. Tassiopoulos. 'Noninferiority Trials'. In: *New England Journal of Medicine* 378.3 (Jan. 2018), pp. 303–305. ISSN: 0028-4793. DOI: 10.1056/NEJMc1714668. URL: http://www.nejm.org/doi/10.1056/NEJMc1714668.

[221] M. J. Tavares Ranzani-Paiva, C. M. Ishikawa, A. C. Das Eiras and V. R. Da Silveira. 'Effects of an experimental challenge with Mycobacterium marinum on the blood parameters of Nile Tilapia,

Oreochromis niloticus (Linnaeus, 1757)'. In: *Brazilian Archives of Biology and Technology* 47.6 (2004), pp. 945–953. ISSN: 15168913.

[222] E. A. Townsend, V. M. Miller and Y. S. Prakash. 'Sex differences and sex steroids in lung health and disease'. In: *Endocrine Reviews* 33.1 (2012), pp. 1–47. ISSN: 0163769X. DOI: 10.1210/er.2010-0031.

[223] P. Truran. 'Models: Useful but Not True'. In: *Practical Applications of the Philosophy of Science*. SpringerBriefs in Philosophy Giere 1999. Heidelberg: Springer International Publishing, 2013, pp. 61–67. ISBN: 978-3-319-00451-8. DOI: 10.1007/978-3-319-00452-5_10.

[224] A. F. Valledor, M. Comalada, L. F. Santamaría-Babi, J. Lloberas and A. Celada. 'Macrophage Proinflammatory Activation and Deactivation. A Question of Balance'. In: *Advances in Immunology* 108.C (2010), pp. 1–20. ISSN: 15578445. DOI: 10.1016/B978-0-12-380995-7.00001-X.

[225] I. Vergne, J. Chua, S. B. Singh and V. Deretic. 'Cell biology of mycobacterium tuberculosis phagosome.' In: *Annual review of cell and developmental biology* 20 (2004), pp. 367–394. ISSN: 1081-0706. DOI: 10.1146/annurev.cellbio.20.010403.114015.

[226] K. Y. Vermaelen, I. Carro-Muino, B. N. Lambrecht and R. a. Pauwels. 'Specific migratory dendritic cells rapidly transport antigen from the airways to the thoracic lymph nodes.' In: *The Journal of experimental medicine* 193.1 (2001), pp. 51–60. ISSN: 0022-1007. DOI: 10.1084/jem.193.1.51.

[227] A. J. Verrall, M. G. Netea, B. Alisjahbana, P. C. Hill and R. van Crevel. 'Early clearance of Mycobacterium tuberculosis: a new frontier in prevention'. In: *Immunology* 141.4 (2014), pp. 506–513. ISSN: 00192805. DOI: 10.1111/imm.12223.

[228] C. L. Vestergaard and M. Génois. 'Temporal Gillespie Algorithm: Fast Simulation of Contagion Processes on Time-Varying Networks'. In: *PLOS Computational Biology* 11.10 (Oct. 2015). Ed. by

M. Salathé, e1004579. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004579. arXiv: arXiv:1504.01298v1.

[229] H. Waaler, A. Geser and S. Andersen. 'The use of mathematical models in the study of the epidemiology of tuberculosis'. In: *American Journal of Public Health and the Nations Health* 52.6 (1962), pp. 1002–1013. ISSN: 0002-9572. DOI: 10.2105/AJPH.52.6.1002.

[230] R. M. Warren, T. C. Victor, E. M. Streicher, M. Richardson, N. Beyers, N. C. Gey Van Pittius and P. D. Van Helden. 'Patients with Active Tuberculosis often Have Different Strains in the Same Sputum Specimen'. In: *American Journal of Respiratory and Critical Care Medicine* 169.5 (2004), pp. 610–614. ISSN: 00030805. DOI: 10.1164/rccm.200305-714oc.

[231] H. C. Warsinske, R. M. DiFazio, J. J. Linderman, J. L. Flynn and D. E. Kirschner. 'Identifying mechanisms driving formation of granuloma-associated fibrosis during Mycobacterium tuberculosis infection'. In: *Journal of Theoretical Biology* 429 (Sept. 2017), pp. 1–17. ISSN: 00225193. DOI: 10.1016/j.jtbi.2017.06.017.

[232] L. G. Wayne and L. G. Hayes. 'An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence'. In: *Infection and Immunity* 64.6 (1996), pp. 2062–2069. ISSN: 00199567.

[233] J. B. West, C. T. Dollery and a. Naimark. 'Distribution of Blood Flow in Isolated Lung; Relation To Vascular and Alveolar Pressures.' In: *Journal of applied physiology (Bethesda, Md. : 1985)* 19 (1964), pp. 713–724. ISSN: 0021-8987.

[234] J. B. West. 'Blood Flow And Metabolism'. In: *Respiratory Physiology: The Essentials*. 7th. Baltimore, Maryland, USA: Lippincott Williams & Wilkins, 2005. Chap. 4, pp. 35–54.

[235] J. B. West. 'Respiratory System Under Stress'. In: *Respiratory Physiology: The Essentials*. 7th. Baltimore, Maryland, USA: Lippincott Williams & Wilkins, 2005. Chap. 9, pp. 137–152.

[236] J. B. West. 'Ventilation'. In: *Respiratory Physiology: The Essentials*. 7th. Baltimore, Maryland, USA: Lippincott Williams & Wilkins, 2005. Chap. 2, pp. 13–23.

[237] J. B. West. 'Ventilation-Perfusion Relationships'. In: *Respiratory Physiology: The Essentials*. 7th. Baltimore, Maryland, USA: Lippincott Williams & Wilkins, 2005. Chap. 5, pp. 55–74.

[238] *WHO HIV data*. URL: http://apps.who.int/gho/data/view.main.22100?lang=en (visited on 04/04/2019).

[239] *WHO Tuberculosis data*. URL: https://www.who.int/tb/data/en/ (visited on 26/03/2019).

[240] J. E. Wigginton and D. Kirschner. 'A model to predict cell-mediated immune regulatory mechanisms during human infection with Mycobacterium tuberculosis.' In: *Journal of Immunology* 166.3 (2001), pp. 1951–67. ISSN: 0022-1767. DOI: 10.4049/jimmunol.166.3.1951.

[241] A. J. Wolf, B. Linas, G. J. Trevejo-Nunez, E. Kincaid, T. Tamura, K. Takatsu and J. D. Ernst. 'Mycobacterium tuberculosis infects dendritic cells with high frequency and impairs their function in vivo'. In: *J Immunol* 179.4 (2007), pp. 2509–2519. ISSN: 0022-1767. DOI: 179/4/2509[pii].

[242] A. J. Wolf, L. Desvignes, B. Linas, N. Banaiee, T. Tamura, K. Takatsu and J. D. Ernst. 'Initiation of the adaptive immune response to Mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs.' In: *The Journal of experimental medicine* 205.1 (2008), pp. 105–15. ISSN: 1540-9538. DOI: 10.1084/jem.20071367.

[243] S. Wolfram. 'Mechanisms in Programs and Nature'. In: *A New Kind of Science*. Champaign, IL: Wolfram Media, 2002. Chap. 7, pp. 297–361.

[244] World Health Organization. *Ethambutol efficacy and toxicity: literature review and recommendations for daily and intermittent dosage in children*. Tech. rep. 2006. DOI: WHO/HTM/TB/2006.365.

[245] World Health Organization. *Global Tuberculosis Report 2018*. Geneva, 2018.

[246] World Health Organization. *Treatment of tuberculosis: guidelines*. Tech. rep. 4th ed. Geneva, 2010.

[247] Y. Xia, O. N. Bjornstad and B. T. Grenfell. 'Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics'. In: *The American Naturalist* 164.2 (2004), p. 267. ISSN: 00030147. DOI: 10.2307/3473444.

[248] Y. Yamamura, Y. Ogawa, H. Yamagata and Y. Yamamura. 'Prevention of tuberculous cavity formation by immunosuppressive drugs.' In: *The American review of respiratory disease* 98.4 (1968), pp. 720–723. ISSN: 0003-0805 (Print). DOI: 10.1164/arrd.1968.98.4.720.

[249] W. W. Yew and C. C. Leung. 'Update in Tuberculosis 2006'. In: *American Journal of Respiratory and Critical Care Medicine* 175.6 (Mar. 2007), pp. 541–546. ISSN: 1073-449X. DOI: 10.1164/rccm.200611-1653UP.

[250] M. A. Yoder, G. Lamichhane and W. R. Bishai. 'Cavitary pulmonary tuberculosis: The Holy Grail of disease transmission'. In: *Current Science* 86.1 (2004), pp. 74–81. ISSN: 00113891.

[251] D. B. Young, H. P. Gideon and R. J. Wilkinson. 'Eliminating latent tuberculosis'. In: *Trends in Microbiology* 17.5 (2009), pp. 183–188. ISSN: 0966842X. DOI: 10.1016/j.tim.2009.02.005.

[252] D. Zenner, N. Beer, R. J. Harris, M. C. Lipman, H. R. Stagg and M. J. van der Werf. 'Treatment of Latent Tuberculosis Infection: An Updated Network Meta-analysis'. In: *Annals of Internal Medicine* 167.4 (Aug. 2017), p. 248. ISSN: 0003-4819. DOI: 10.7326/M17-0609. URL: http://annals.org/article.aspx?doi=10.7326/M17-0609.

[253] A. I. Zumla et al. 'New antituberculosis drugs, regimens, and adjunct therapies: Needs, advances, and future prospects'. In: *The Lancet Infectious Diseases* 14.4 (2014), pp. 327–340. ISSN: 14744457. DOI: 10.1016/S1473-3099(13)70328-1.